INTERACTION OF B-DNA WITH MONOVALENT CATIONS: THEORY AND PRACTICE IN X-RAY CRYSTALLOGRAPHY

A Thesis
Presented to
The Academic Faculty

by

Tinoush Moulaei

In Partial Fulfillment
Of the Requirements for the Degree
Doctor of Philosophy in Chemistry

Georgia Institute of Technology
December 2004

INTERACTION OF B-DNA WITH MONOVALENT CATIONS: THEORY AND PRACTICE IN X-RAY CRYSTALLOGRAPHY

Approved:

Loren D. Williams, Chairman
School of Chemisty and Biochemistry

Roger Wartell
School of Biology

Donald F. Doyle
School of Chemisty and Biochemistry

Angus P. Wilkinson
School of Chemisty and Biochemistry

Nicholas V. Hud
School of Chemisty and Biochemistry

Date Approved: November 19th, 2004

DEDICATION

A small token dedicated to Iranians and Americans with the hope that one day all of you experience true democracy and freedom.

ACKNOWLEDGEMENTS

I would like to acknowledge and thank my advisor, Dr. Loren D. Williams for his guidance through the years that I have worked with him. He has been a great educator, always providing his expertise and views on the problem at hand to propel me forward and yet keeping the mystery that is needed to motivate me to solve the problem. He also provided the financial support for my studies. My family, both in Iran and in the US, has always supported me financially and emotionally throughout my education. Not only am I thankful for their support, but I must also commend them for enduring my stubborn character all my life and especially over my 14 years at Georgia Tech! Also, Dr. Stephen Quirk was my previous advisor and was instrumental in my acceptance into the program and instructing me in the early part of my career.

I would like to thank other members of the Williams' group that have worked with me throughout the many years, Tatsuya Maehigashi, Derrick Watkins, Dr. Seiji Komeda, Dr. Eli Hershkowitz, Srividia Mohan, Dr. Mary Peeks, Lopa Bhatt, Sarah Tannenbaum, Audra Morabito, Dr. Kristen Woods, Dr. Shelley Howerton, Dr. Ryan Burnett, Dr. Denise Jones, Dr. Marc Vogt, Brian Lynch, Tanya Civijan, Melissa Kerzic, Amy Vinson, Jill Heidinger, Jen Leader, Chad Sines, and Jennifer Dunn. I have had the priviledge of working with many gifted and kind individuals.

The two final projects in this thesis benefited from the expertise of three collaborators, whom I must thank. Dr. Barry Gold kindly provided the modified DNA oligomers for these projects and also provided advice in the preparation of the manuscript. Dr. Micheal Stone provided NMR data, and also assisted with corrections to

the manuscript. Dr. Luis Markey provided his expertise in interaction of cations with these modified DNA molecules.

I would also like to thank all faculty members that have consistently provided me with advice. There are many, but I must mention my committee members Dr. Donald Doyle, Dr. Nicholas Hud, Dr. Angus Wilkinson, and Dr. Roger Wartell. I would also like to thank other members of the Tannenbaum family who have very been supportive, Dr. Allen Tannenbaum, Dr. Rina Tannenbaum, and Dr. Manny Tannenbaum. Sarah Tannenbaum provided assistance for parts of this work and has been an invaluable friend. Dr. Allen Orville, and his lab members, George Lountos, Akanksha Nagpal, and Linda Manning have been supportive in many ways, sharing ideas, and providing me with their knowledge of crystallography.

Derrick Watkins, Tatsuya Maehigashi, Dr. Seiji Komeda, Sarah Tannenbaum, George Lountos, and Dr. Kristen Woods deserve special recognition for collecting the synchrotron data sets used in this thesis. Last but not least, I would like to thank Dr. Thomas Turnabene for his guidance in my many years at Georgia Tech. I consistently sought and received his advice in making many crucial academic decisions.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

This work attempts to address some fundamental questions about the interaction of B-DNA and monovalent cations. The sequence dependence of DNA conformation has been demonstrated and experiments supporting its veracity have been acknowledged and verified. However, it remains to be seen what role, if any, cations play in this process. A plethora of experimental evidence has established that monovalent cations do localize around DNA, and especially in the major and minor grooves. This thesis will concentrate on three aspects of the current theories on the interaction of B-DNA with monovalent cations.

The first section examines the interpretation given to the isolated electron density peaks in the solvent/counter-ion region of nucleic acids x-ray crystal structures. In the solvent/counter-ion region, distinction between atoms and molecules with similar number of electrons and geometry is made difficult by the very nature of the x-ray diffraction. The diffraction of x-rays is mainly affected by the density of electrons and not by the nuclear structure (i.e. identity of the atom). The process of fitting the model to the available diffraction data can compromise the crystallographer's subjectivity, especially in fitting the solvent/counter-ion region of the model. The density maps that are used in the process are based on the scattering power of the available electrons at different regions of the crystal. As a result, in the solvent/counter-ion region of nucleic acids structures the atomic assignments that are made are less objective and more based on preference or established dogma.

The data presented show that electron density peaks assigned to water molecules can as easily be assigned to cationic species with similar number of electrons. This hypothesis is demonstrated by substituting fully occupied sodium and partially occupied potassium cations for the water molecules in two B-DNA structures. The occupancy values are also shown to be arbitrary to some degree, since the refinement program may adjust other parameters to compensate for over or under occupied regions. All the experiments show that within limits, there is minimal penalty in the refinement statistics as a result of changing either the identity or the occupancy of the peaks in the counterion region.

Next, a comparative study of two B-DNA structures outlines the differences in the interaction of B-DNA with various monovalent cations. Specific emphasis is placed on rubidium (I) and thallium (I), two monovalent heavy atoms, which are often used to probe for potential cation binding sites in macromolecular crystallography. The initial assumption that thallium behaves in a manner similar to the group I metals proved to be an oversimplification. Although thallium (I) shares many characteristics with group I metals especially potassium, there are differences arising from its electronic structure, which need to be considered in molecular structures. This experiment also demonstrates the reproducibility and reliability of anomalous scattering as a tool for locating potential rubidium and thallium binding sites.

Finally, the effects of cations tethered into the major groove of B-DNA at the A-tracts are explored. The tethered cation is an 3-aminopropyl moiety covalently linked to the C5 of deoxyuridine. Two x-ray crystal structures are solved by molecular replacement and analyzed. Axial bending of the DNA is observed in one structure

containing a total of four tethered cations. Along with DNA bending, there is also a lack of monovalent cations in the major groove as compared to a B-DNA structure without the cationic tether. These results suggest a possible mechanism for DNA bending by charged side chains of proteins and binding of free cations in the major groove. The other structure in this series contains a total of two tethered cations and exhibits no bending. Comparison of both structures shows that crystal-packing forces interfere with deformation of the DNA backbone. The result of this interference is that in one structure the DNA does not bend at all, while in the second structure the DNA is partially bent. The bending observed is not to the degree suggested by the published solution NMR studies of the same DNA sequence.

CHAPTER I

BACKGROUND

The Bending of the Helix

In 1953 a new chapter in the history of structural biology was opened. The culmination of research done by many scientists including Francis Crick, Rosalind Franklin, James Watson, and Maurice Wilkins resulted in the discovery of the structure of DNA (Franklin & Gosling, 1953; Watson & Crick, 1953). Since then, the scientific community has been focusing on ever-finer details of this remarkable molecule. The original model of DNA published in 1953 has been expanded to include A-, B-, B*- and Z-forms, and has been shown to possess a surprising degree of elasticity.

In solution, DNA has a persistence length of about 150 bp (Hagerman, 1988), yet a 146 bp stretch of DNA makes almost two turns around an octamer of histone proteins (Luger et al., 1997). Other proteins also have the ability to induce extreme axial bending in DNA. The catabolite activator protein (CAP) induces almost 90° of bending around itself upon binding to its target site on DNA (Kahn & Crothers, 1998; Passner & Steitz, 1997). Remarkably, proteins can also induce DNA bending away from them. The TATA-element binding protein (TBP) binds into the minor groove of DNA at a specific sequence and induces an almost 90° bend away from itself (Nikolov et al., 1996; Nikolov et al., 1995). The zinc finger (Ikeda, Nagano & Kawakami, 1993) and bZIP (Kerppola & Curran, 1991; Paolella et al., 1997) families of transcription factors also induce DNA bending.

There is a long list of proteins that induce bending in DNA. One common theme among this class of proteins appears to be gene activation and repression. The perturbation of the DNA model published in 1953 is one prominent mode of exerting genetic control employed by cells. It is intuitive to expect the possibility of DNA bending at the hands of certain proteins. Less intuitive is the deformation of DNA based on its sequence especially when considering the interactions between DNA and the charged species in its milieu, a point that will be expounded in this thesis.

This thesis is concerned with three aspects of x-ray crystallography and B-DNA conformation. In the first section, the ambiguity inherent to interpreting the counterion region of x-ray crystal structures is considered. The experiments focus on examination of the subjectivity required to interpret the solvent species in crystallography and special attention is given to two published structures of the Dickerson-Drew dodecamer $[d(CGCGAATTCGCG)]_2$ (hereon referred to as DDD). In the next chapter, two structures of DDD molecule grown in presence of Rb^+ and Tl^+ are compared. The goal of this section is to compare and contrast these two monovalent cations to other group I monovalent cations. Finally, two structures of B-DNA with monovalent cations covalently linked to the major groove are analyzed. The impact of these tethered cations on the overall DNA structure and the distribution of free cations is studied.

Sequence Dependence of DNA Conformation

Many studies have probed the variations in the structural properties of double-stranded DNA based on its sequence. One of the earliest discoveries was DNA bending at A-tracts (Koo, Wu & Crothers, 1986; Selsing et al., 1979). One A-tract can bend the

DNA axis by 17°-21° (Koo et al., 1990). Axial bending has also been observed in consecutive runs of guanines and cytosines, G-tracts (Brukner et al., 1994; Goodsell et al., 1993). More local sequence dependent effects have been seen in X-ray crystal structures of short DNA molecules. These include variations in basepair parameters such as rise, roll, and propeller twist (Figure 1) in addition to deviation from ideality of groove parameters (Drew et al., 1981; Shui et al., 1998a).

Sequence dependence of DNA conformation is thought to have an effect on DNA-protein interactions and ultimately genetic control in biological systems. The serum response factor appears to initiate transcription by recognizing and binding to curved DNA and sequestering other proteins (Pellegrini, Tan & Richmond, 1995). The 434 repressor and regulatory protein Cro bind to six naturally occurring operator sequences with different affinities. Central non-contacting bases in the operators modulate the affinity of both of these proteins towards the operator sequences (Koudelka et al., 1988; Koudelka, Harrison & Ptashne, 1987). TBP binds to its target site and bends the DNA without making specific contacts with DNA at this site. It is plausible that TBP recognizes its target site through the sequence dependent deviations in the DNA structure. In contrast, most proteins that bind specific DNA sequences recognize their target site by making direct contact with DNA bases and detecting the hydrogen bonding of the functional groups in the major and minor grooves (Seeman, Rosenberg & Rich, 1976).

It is established that the chemical properties of the nucleic acid bases dictate the conformation of the DNA. Among these chemical properties, the hydrophobic stacking of the bases and the hydrogen bonding of purines with pyrmidines are of greatest

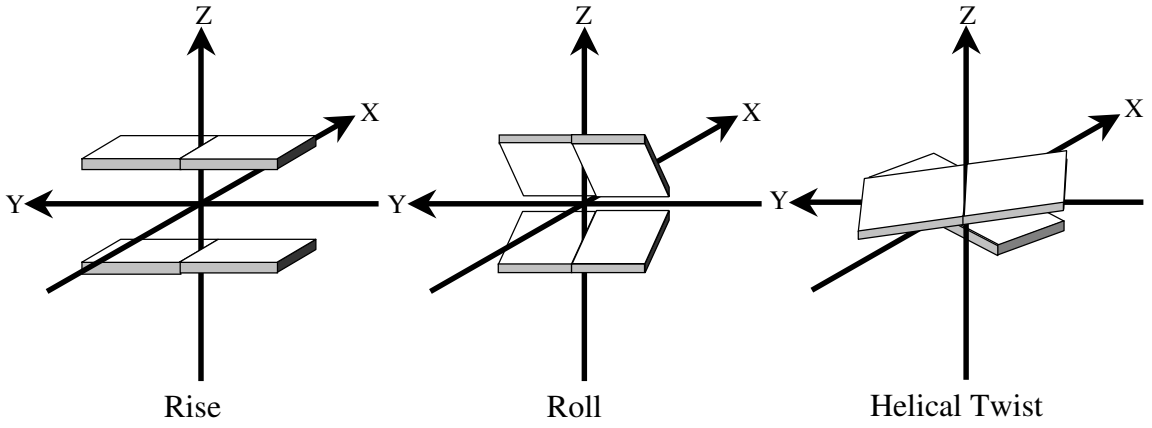

Figure 1. Schematic drawing of inter-basepair parameters rise, roll, and helical twist.

importance. DNA adapts conformations that maximize the stacking interaction hence excluding water molecules from contact with the hydrophobic faces of the nucleic acid bases. In addition, the orientation that the bases adapt with respect to each other allows for the strongest possible hydrogen bonding. Finally, the double helix is ultimately at the mercy of the coulombic forces repelling the negatively charged backbones of each strand. Therefore, in the absence of positively charged species to neutralize the phosphate groups, the backbones cannot approach each other.

There are two views on the sequence dependence of DNA conformation. Both views accept that local conformation of DNA is sequence dependent and leads to global changes in the DNA structure. One set of models attribute the sequence dependent local conformation of DNA to hydrophobic forces, hydrogen bonding, and phosphate-phosphate repulsion discussed above. Water molecules play important roles and exert their effect through hydrogen bonding with the peripheral groups on the bases and the backbone (Allemann & Egli, 1997; Fratini et al., 1982; Grzeskowiak et al., 1991). In these models, even transition from B- to A-form of the duplex $[d(GGGCCC)]_2$ under

increasing cation concentration has been attributed to the reorganization of water molecules (Ng & Dickerson, 2002).

The alternative view is that DNA sequesters cations into its minor and major grooves in a sequence dependent manner. The effects of cations on the electrostatic interactions of DNA bases with each other and their environment determines the local conformation of DNA (Shui et al., 1998a; Shui et al., 1998b). It is difficult to locate cations in x-ray crystal structures and NMR solution structures, and to verify the accuracy of either mechanism for the sequence dependence of the structure of DNA. What is apparent is a charge imbalance in almost all nucleic acid structures. There simply is not enough positive charge to neutralize the amount of negative charge on the backbone of DNA in the available structures.

Classically, this discrepancy has been explained by the delocalized counterions around DNA in accordance with the polyelectrolyte theory (Anderson & Record, 1995; Manning, 1978; Record, Anderson & Lohman, 1978). The polyelectrolyte theory proposes that the concentration of the cations in the counterion condensation layer immediately adjacent to the nucleic acid molecules is independent of the ionic strength of the solution. These cations are delocalized, always hydrated and do not make direct contact with the DNA, but do neutralize the negative charge of the phosphate groups of the backbone (Manning, 1978). However, much of the recent evidence suggests that cations do localize in a sequence specific manner around DNA, and especially in the minor and major grooves.

A-tracts, G-tracts and Generic DNA

Among the sequences that impart a specific conformation to DNA, A-tracts and G-tracts are distinctive, because of the pronounced effects that they induce and the amount of available data. Hud and Plavec define A-tracts as four or more basepairs long with only ApA or ApT base-steps and no TpA base-steps (Hud & Plavec, 2003). Sequences containing TpA-steps do not induce axial benging in DNA (Hagerman, 1986; Hagerman, 1990; Stefl et al., 2004). Therefore, A-tracts may contain ApT base-steps but not TpA base-steps. Such precise definition arises from their properties and the enormous volume of available data on sequences that share similar characteristics and can be defined as such. There is less data available on G-tracts and as a result G-tracts are less precisely defined. G-tracts contain consecutive runs of guanines and cytosines with one or more GG steps. Unlike A-tracts, both GC and CG steps are allowed in G-tract sequences. All other DNA sequences that are not qualified by these two sets of criteria are defined as generic DNA (Hud & Plavec, 2003).

A-tract DNA is commonly classified as B*-form DNA, with a narrower minor groove than B-form DNA, high propeller twist, and the tendency to bend DNA (Alexeev, Lipanov & Skuratovskii, 1987; Nelson et al., 1987). Gel mobility shift assays have shown that naturally occurring phased A-tracts can induce DNA bending. In mitochondrial kinetoplast DNA, multiple A-tracts are separated by intervening sequences such that the minor groove of the A-tracts face the same direction. DNA is bent successively at each A-tract and the cumulative result is large overall curvature of the entire molecule. The curved DNA migrates at the same rate as a linear DNA of over twice its length (Marini et al., 1982).

G-tracts undergo a transition from B- to A-form DNA with increasing cation concentration. They are specially susceptible to the increased concentration of divalent cations (Wolk et al., 1989; Xu, Jampani & Braunlin, 1993). G-tracts also bend DNA, although that requires multivalent cations (Brukner et al., 1994). A-tracts bend DNA towards the minor groove (Zinkel & Crothers, 1987), while G-tracts bend the DNA towards the major groove (Brukner et al., 1994; Dlakic & Harrington, 1995). The structure of generic DNA is described as the canonical B-form DNA. Examples of A-, B-, and B*-forms of DNA are shown in Figure 2.

There is a large body of evidence that suggests both A- and G-tract DNA sequester cations. Monovalent cations localize in the minor groove of A-tracts both on the floor and at top of the groove (Howerton et al., 2001; Hud, Sklenar & Feigon, 1999; Shui et al., 1998a). Divalent cations also associate at the mouth of the A-tract minor groove in association with phosphates (Minasov, Tereshko & Egli, 1999; Sines et al., 2000).

There is also an absence of cations from the major groove of A-tracts (Howerton et al., 2001; Hud & Feigon, 2002). In contrast, G-tracts tend to localize cations in the major groove and not the minor groove (Xu, Deng & Braunlin, 1993; Xu, Shoemaker & Braunlin, 1993). The degree of axial bending induced by G-tracts is dependent on the concentration of divalent cations (Brukner et al., 1994). The available x-ray crystal structures show that there is a range of conformations that G-tracts will adopt. At the extremes of this spectrum of conformations are the B- and A-form DNA. Where the structure of a G-tract will be in this spectrum depends on the concentration of divalent cations in the major groove (Hud & Plavec, 2003; Lu, Shakked & Olson, 2000). This

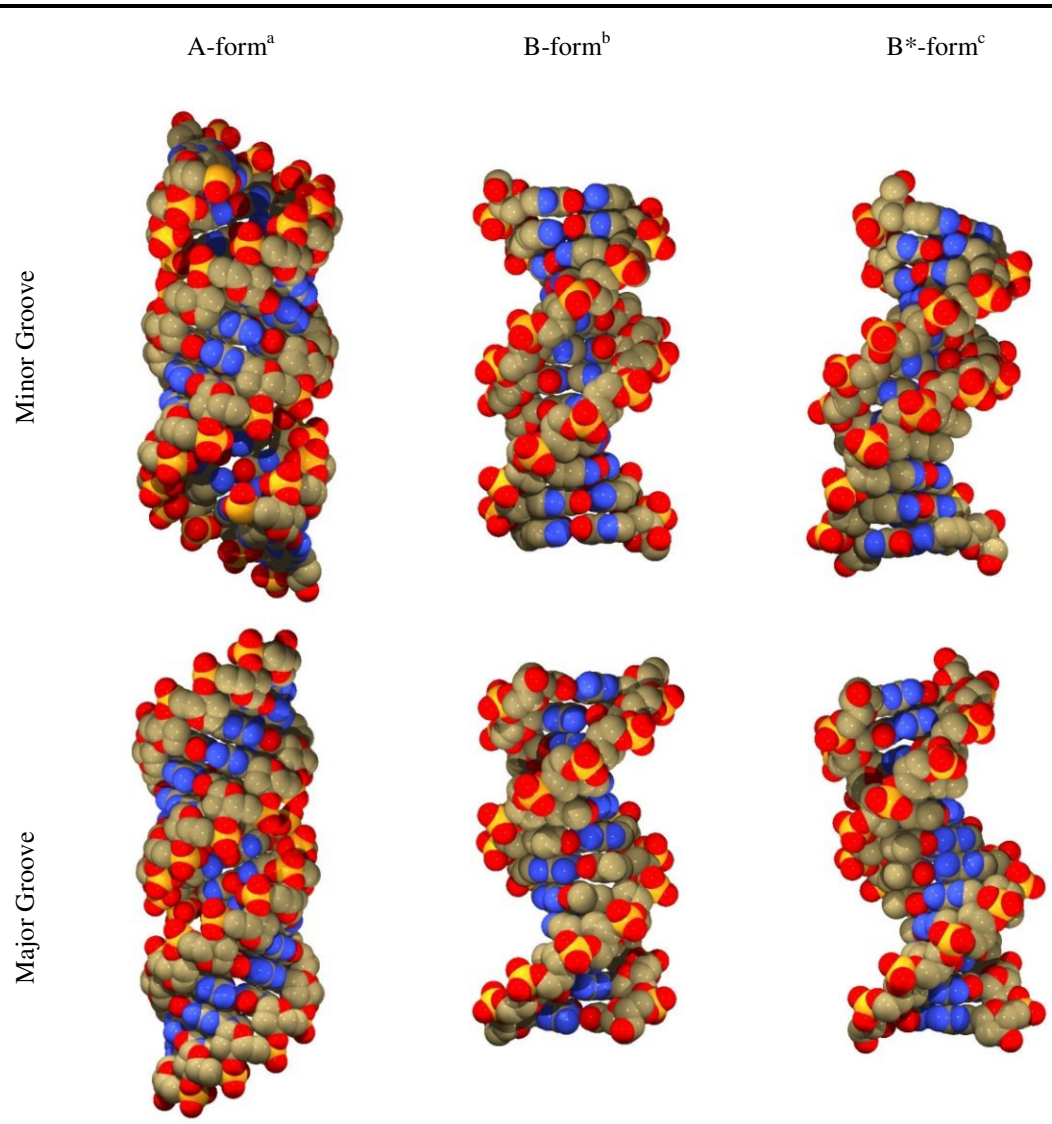

Figure 2. A-, B-, and B*-forms of DNA.
[a]Ideal structure of A-form DNA (left). The sequence is $[d(CGCGCGCGCGCGCGCG)]_2$. [b]Ideal B-form structure of DDD (center). [c]X-ray crystal structure of a poly(dA)·poly(dT) tract (right) (Nelson et al., 1987). Coloring scheme: carbon – grey, oxygen – red, nitrogen – blue, phosphorous – yellow.

propensity for B- to A-form transition is lacking in A-tracts. Even at very low water activity, A-tracts resist adopting the A-form conformation (Leslie et al., 1980).

There are many models for the mechanism of DNA bending by non-generic DNA sequences. They differ mainly on the location of the axial bending. In general, they can

be placed in two classes, the wedge and the junction classes. The wedge class of models posits that there is a global axial bending distributed smoothly through the A-tract and G-tract (non-generic DNA) sequences. Bending occurs as a result of 'wedges' spaced in between the stacked basepairs. These wedges are the conceptualized effect of variations in roll and tilt parameters between consecutive basepairs (Trifonov, 1991; Ulanovsky & Trifonov, 1987). The junction class models propose that the bending occurs at the junctions between the A-tracts and G-tracts, and generic DNA (Koo et al., 1986). A model put forth by Dickerson and co-workers proposes that A-tracts are straight and sequences rich in guanines and cytosines are curved. In this model, DNA curvature is independent of cations (Goodsell et al., 1993). Crother and co-workers have resolved the major differences between the wedge and junction models (Haran, Kahn & Crothers, 1994). It is now well accepted that DNA bending occurs at the junctions between A- and B*-form DNA with B-form DNA. What remains to be conclusively demonstrated is the mechanism of the bending process.

Mechanism of DNA Bending: Coarse and Fine Controls

Because of the polyanionic nature of DNA, significant bending of its axis is energetically unfavorable. The DNA resistance to bending arises from the repulsive forces between the negatively charged phosphate groups. Bending of DNA would require a closer approach between some phosphates, and would increase these repulsive forces. Hence, DNA is naturally rigid and inflexible (Olson et al., 1998). Alternatively, Hagerman and Hagerman propose that the intrinsic rigidity of the duplex arises from resistance to disruption of base-stacking interaction. These stacking interactions

thermodynamically drive the duplex towards a rigid structure, in which stacking is maximized. In this view, phosphate-phosphate interactions are salt and sequence dependent and play a minor role (Hagerman & Hagerman, 1996). One interpretation of these models would be that base-stacking interactions are the major determinant and make coarse adjustments to rigidity (and global structure), while the phosphate-phosphate repulsive and sequence-dependent effects cause fine adjustments on DNA rigidity (and local structure).

One of the earliest models for describing the bending of DNA around proteins proposes that asymmetric neutralization of the phosphodiester charge along one face of the helix is required for the axial bending to become energetically favorable. This site-specific charge neutralization is achieved when regions with positive electrostatic potential on the protein interact with the DNA backbone. The result of charge neutralization is the electrostatic collapse of the backbone and axial bending (Manning et al., 1989; Mirzabekov & Rich, 1979). In fact, when neutral methylphosphonate residues are phased within guanine and cytosine rich sequences, DNA bending is observed at the location of the neutral residues (Strauss & Maher, 1994). Additionally, a host of tethered cations attached to pyrimidine bases have been used to mimic the electrostatic effect of protein binding to DNA. These cations have been covalently linked to the C5 of pyrimidines such that they are positioned in the major groove of the duplex. Phased patches containing these zwitterionic residues were incorporated into longer DNA molecules. The result was that bending was observed at the location of the modified bases and towards the face containing the modifications (Gold, 2002; Heystek et al., 1998; Liang et al., 1995; Strauss et al., 1996a; Strauss et al., 1996b).

Maher and co-workers have also constructed mutants of GCN4, which induce DNA bending, *in vitro*. GCN4 is a member of bZIP family of DNA binding proteins. Experiments have shown that although GCN4 binds to its target DNA, it does not induce bending. Substitution of three neutral amino acids with basic residues enables GCN4 mutants to bind and bend DNA (Strauss-Soukup & Maher, 1997). These experiments support the electrostatic collapse model for DNA bending by proteins.

Rouzina and Bloomfield propose a mechanism in which the multivalent cationic species in the groove repel the cations associated with the phosphates. The unscreened phosphates can then collapse towards the cationic species, leading to narrowing of the groove, and ultimately to axial bending of DNA (Rouzina & Bloomfield, 1998). Hud and Plavec have pointed out that these models may not sufficiently describe the bending of DNA as a result of cation binding into the major and minor grooves (Hud & Plavec, 2003). The electrostatic collapse model also does not satisfy results obtained by Maher and co-workers in experiments done as an extension of their work with neutral methylphosphonate residues. The original experiment contained a racemic mixture of methylphosphonates. When they used residues with only (R_p)-methylphosphonate substitutions in their experiments, the degree of observed DNA bending was reduced (Strauss-Soukup et al., 1997).

An alternative model proposed by Hud and coworkers describes the minor and major grooves as "flexible ionophores" engaged in a "tug of war" to sequester cations and affect the conformation of DNA. Hence, the binding of the cations in the grooves is the determining factor in its conformation. A-tracts and G-tracts present cases where one groove has a clear victory over the other (Hud & Plavec, 2003; Hud & Polak, 2001).

Table 1. Survey of the current structures in the NDB with resolution below 1.0 Å.

ID	Description	Res. (Å)[a]	*R*	H_2O[b]	PC[c,b]	Mg^{2+} [d]	Other[d]	Δq[e]
AD0007[f]	A-DNA Decamer	0.83	11.5	177	2	1.5	0.0	-11.0
BD0023[g]	B-DNA Decamer	0.74	10.5	238	0	0.0	5.1(Ca^{2+})	-10.2
BD0033[h]	B-DNA Decamer	0.98	14.1	110	0	6.3	0.0	-5.4
BD0034[h]	B-DNA Decamer	0.98	14.1	116	0	0.0	4.5($Ca2^{+}$)	-9.0
BD0035[h]	B-DNA Decamer	0.98	14.1	99	0	7.8	0.0	-2.4
BD0036[h]	B-DNA Decamer	1.00	12.1	140	0	0.0	6.0(Ca^{2+})	-6.0
BD0037[i]	B-DNA Nonamer	0.89	13.5	151	0	6.8	1.0(Cl^{-})	-3.5
DD0027[j]	Z-DNA Hexamer	0.90	15.2	101	0	1.0	0.0	-9.0
DDF027[k]	Z-DNA Hexamer	1.00	16.1	61	1	3.0	1.0(Na^{+})	0.0
ZD0003[l]	Z-DNA Hexamer	1.00	16.5	88	2	1.0	0.0	0.0
ZD0004[m]	Z-DNA Hexamer	0.60	16.0	40	1	0.0	0.0	-6.0
ZD0005[n]	Z-DNA Hexamer	0.95	8.7	83	1	0.7	0.0	-5.3
ZDF001[o]	Z-DNA Hexamer	0.90	14.0	74	2	1.0	0.0	0.0
ZDF002[p]	Z-DNA Hexamer	1.00	17.5	84	0	4.0	0.0	-2.0
ZDF029[q]	Z-DNA Hexamer	1.00	18.5	47	1	0.0	0.0	-6.0
ZDF035[r]	Z-DNA Hexamer	1.00	18.0	62	2	0.0	2.0(Na^{+})	0.0
ZDF052[s]	Z-DNA Hexamer	1.00	19.1	74	1	3.0	1.0(Na^{+})	0.0
ZDF053[t]	Z-DNA Hexamer	1.00	19.0	63	2	2.0	0.0	+2.0
AR0019[u]	Hybrid Nonamer	0.97	14.8	84	0	0.0	0.3(SO_4^{2-})	-16.3
UD0041[v]	DNA Quadruplex	1.00	14.9	180	0	0.0	1.0(V^{4+})	-20.0
UDF062[w]	DNA Quadruplex	0.95	15.2	140	0	0.0	3.5(Na^{+}),2.3(Ca^{2+})	-12.0
UR0014[x]	RNA Quarduplex	0.61	11.2	106	1	0.0	2.0(Sr^{2+}),3.0(Ca^{2+}),2.0(Na^{+})	-6.0
			Total:	2318	16	38.1	34.7	-128.1

[a]Resolution [b]Total number of molecules and atoms regardless of their occupancy [c]Polycation [d]Sum of the occupancies [e]The sum of the charges from all charged species accounting for partial occupancies [f](Egli et al., 1998) [g](Kielkopf et al., 2000) [h](Chiu & Dickerson, 2000) [i](Soler-Lopez, Malinina & Subirana, 2000) [j](Bennett et al., 2000) [k](Ohishi et al., 1991) [l](Ohishi et al. unpublished) [m](Tereshko et al., 2001) [n](Dauter & Adamiak, 2001) [o](Wang et al., 1979) [p](Gessner et al., 1989) [q](Egli et al., 1991) [r](Bancroft et al., 1994) [s](Ohishi et al., 1996a) [t](Ohishi et al., 1996b) [u](Masquida, Sauter & Westhof, 1999) [v](Cardin et al. unpublished) [w](Phillips et al., 1997) [x](Deng, Xiong & Sundaralingam, 2001)

Locating Monovalent Cations in X-ray Crystal Structures

A complete description of the sequence dependence of DNA conformation must include its interaction with the counterions in its milieu. Yet, a survey of the structures available in the Nucleic Data Base (NDB) (Berman, Zardecki & Westbrook, 1998) shows a pronounced lack of counterions and essentially a charge imbalance.

There are not enough cations to neutralize the negative charge of the backbone phosphates. Table 1 lists all of the current x-ray crystal structures in NDB with a resolution better than 1.0 Å. Only two existing structures are not listed in this table. One structure contains heavily modified guanosines (AD0025), and the other has no solvent molecules and atoms listed (ZDF013). Mono-, di-, and trinucleotides are also excluded. Of the 22 structures listed in Table 1, five achieve charge neutrality, all of which are Z-DNA hexamers, and one has a net positive charge of +2! There are only 9 Na^+ ions and over 2318 water molecules, while in most cases, Na^+, K^+, or both are present in the crystallization solution. The sum of the charges in all of the structures is -128. The total charge for B-DNA structures is –36.5 with no monovalent cations present.

Clearly, such a large negative charge is unrealistic. Charge neutrality may also be an unrealistic standard to set for x-ray crystal structures, since many species in the crystal will be disordered. However, such large negative potentials as listed in Table 1 are unattainable. This problem arises from the inability of the commonly used techniques to distinguish cations from water molecules and from each other in most cases. Biologically relevant cations, Na^+, K^+, and Mg^{2+} have similar number of electrons and x-ray scattering power as water molecules (see Table 3, Chapter II). The problem is compounded since in

many cases the cation binding sites are partially occupied. Therefore, a 60% occupied K^+ site is indistinguishable form a fully occupied water molecule.

Although identification of cations is difficult, a large volume of evidence supports the localized binding of cations to nucleic acids. These include x-ray crystallography studies (Howerton, Nagpal & Williams, 2003; Howerton et al., 2001; Tereshko, Minasov & Egli, 1999; Tereshko et al., 2001; Woods et al., 2000), solution NMR experiments (Cesare Marincola, Denisov & Halle, 2004; Denisov & Halle, 2000; Hud et al., 1999), and molecular dynamics simulations(Hamelberg et al., 2000; Hamelberg, Williams & Wilson, 2002; McConnell & Beveridge, 2000; Young, Jayaram & Beveridge, 1997).

One method for identification of cation binding sites, is to replace them with heavier cations possessing stronger scattering power and in many cases anomalous scattering. In this thesis, two commonly used heavy monovalent cations, Rb^+ and Tl^+, will be used to identify cation binding sites in native and modified B-DNA dodecamers. Tereshko and co-workers have used anomalous scattering to identify positions of Rb^+ and Cs^+ in modified A-DNA decamers (Cesare Marincola et al., 2004). Tl^+ has been used as a mimic of K^+ to probe structures of nucleic acids (Basu et al., 2000; Howerton et al., 2003; Howerton et al., 2001) and proteins (Badger et al., 1994a; Badger, Li & Caspar, 1994b; Brown, 1988; Gill & Eisenberg, 2001; Gursky et al., 1992a; Gursky et al., 1992b; Loria & Nowak, 1998; Pedersen et al., 1998; Stroud, Kay & Dickerson, 1974; Villeret et al., 1995). One important question that will be addressed is whether the heavy cation substitutes are good representatives of the chemical and physical properties of the lighter, biologically relevant cations.

CHAPTER II

METHODS

Sample Preparation and Crystallization

The thrust of this work is in the cationic environment of B-from DNA. Therefore, extreme care was taken at all steps to control the cation species introduced in the crystallization solution. All DNA samples were desalted and purified to remove as much of the cations introduced during the synthesis process as possible. The DNA samples were annealed prior to crystallization. The sample was melted at 80°C for 10 minutes and cooled to room temperature over 1 hour. The annealed samples were stored at -20°C. For the two structures containing the Z3dU residue (Figure 3), the cacodylate solution used was buffered to the desired pH using a solution of the base form of spermine. This stock solution contained 1 mole of spermine for every 5 moles of cacodylate (pH 6.5). The same spermine solution was also buffered using acetic acid to prepare a stock solution of spermine acetate (pH 6.5) and was used to add spermine. For the structure containing unmodified native DNA, rubidium cacodylate was prepared by titrating a cacodylic acid solution with rubidium hydroxide.

The crystallization of each sample was initiated with a coarse 2D grid of magnesium acetate and spermine acetate (pH 6.5). Favorable conditions from these crystallization setups were optimized using a finer 2D grid of the same two chemicals. If the goal of the experiment was to crystallize a molecule with heavy atom derivatives (Rb^+ and Tl^+), conditions with the lowest Mg^{2+} and spermine concentration were favored as

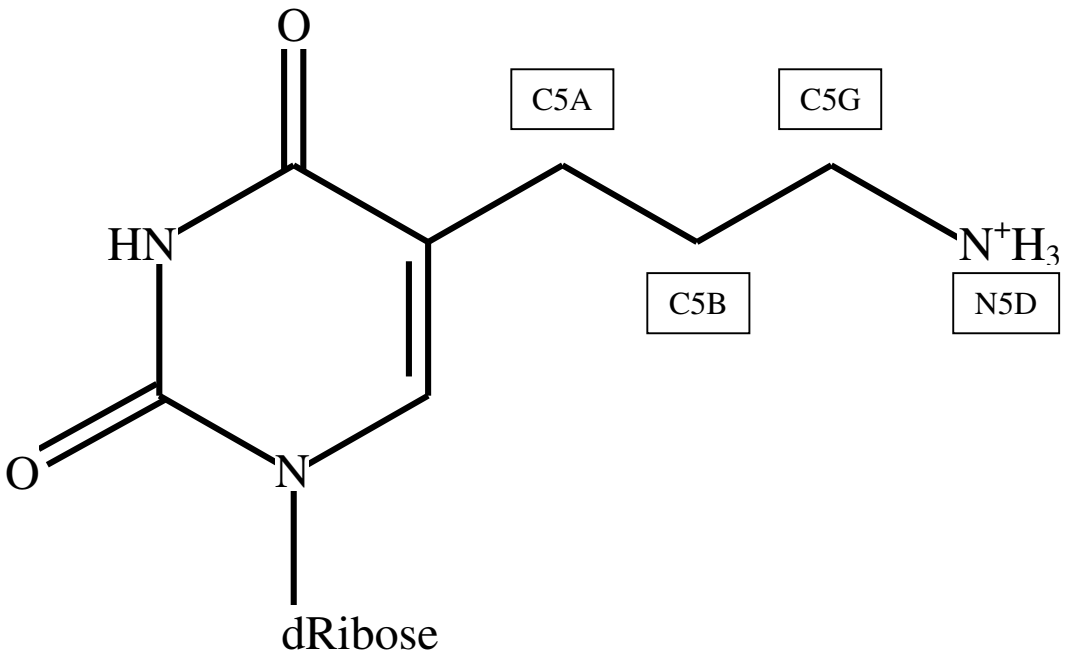

Figure 3. Structure of 5-(3-aminopropyl)-2'-deoxyuridine, Z3dU.
Atom names for the 3-aminopropyl modification are given in the boxes.

long as the quality of the crystal was not compromised. All viscous chemicals such as 2-methyl-2,4-pentanediol (MPD) and detergents were distributed among the different component in the crystallization solution for fast and optimal mixing. For example, it was preferred to use equal volumes of two solutions containing 4% MPD than to use equal volumes of a solution with no MPD and a solution with 8% MPD. The main reason for this approach was to insure reproducibility. Crystals were grown under both sitting drop and hanging drop conditions. Standard biochemical and crystallographic techniques were followed for other details of sample preparation and crystal growth (Howerton, 2002).

Data Collection and Processing

The data for the Rb^+ form of DDD were collected on an in-house Rigaku/MSC rotating anode generator with Osmic blue confocal mirrors (model RU-H3R) and a Raxis IV^{++} image plate detector. This data set was indexed and integrated using MOSFLM version 6.2.2 and scaled using SCALA. Dr. Shelley Howerton performed this work using standard techniques.

The data for the structures containing the modified residue Z3dU were collected at beamline 22-ID in the SER-CAT facilities of Advanced Photon Source (APS) at Argonne National Laboratory near Chicago, Illinois. The data were collected on a MAR165 detector and processed using the HKL2000 version 1.97.7 software package. Because of certain US government policies this author was unable to travel to synchrotron radiation facilities. Hence, other group members and colleagues were kind enough to collect the synchrotron data for these experiments. Their contribution and gracious assistance to this thesis has been noted in the acknowledgements. As a result of this, none of the anomalous data sets were collected at the optimal wavelength to maximize the anomalous signal.

The crystals of the dodecamers containing Z3dU were twinned. Initially the data were integrated, reduced and scaled at APS. These reduced data sets were of poor quality and electron density maps calculated using these data resulted in missing sum (2|Fo|-|Fc|) density throughout the structures (Figure 4). The images collected from these crystals were examined in detail and the refinement parameters were carefully monitored to exclude the minimum number of twinned frames. In general these frames have mosaicity

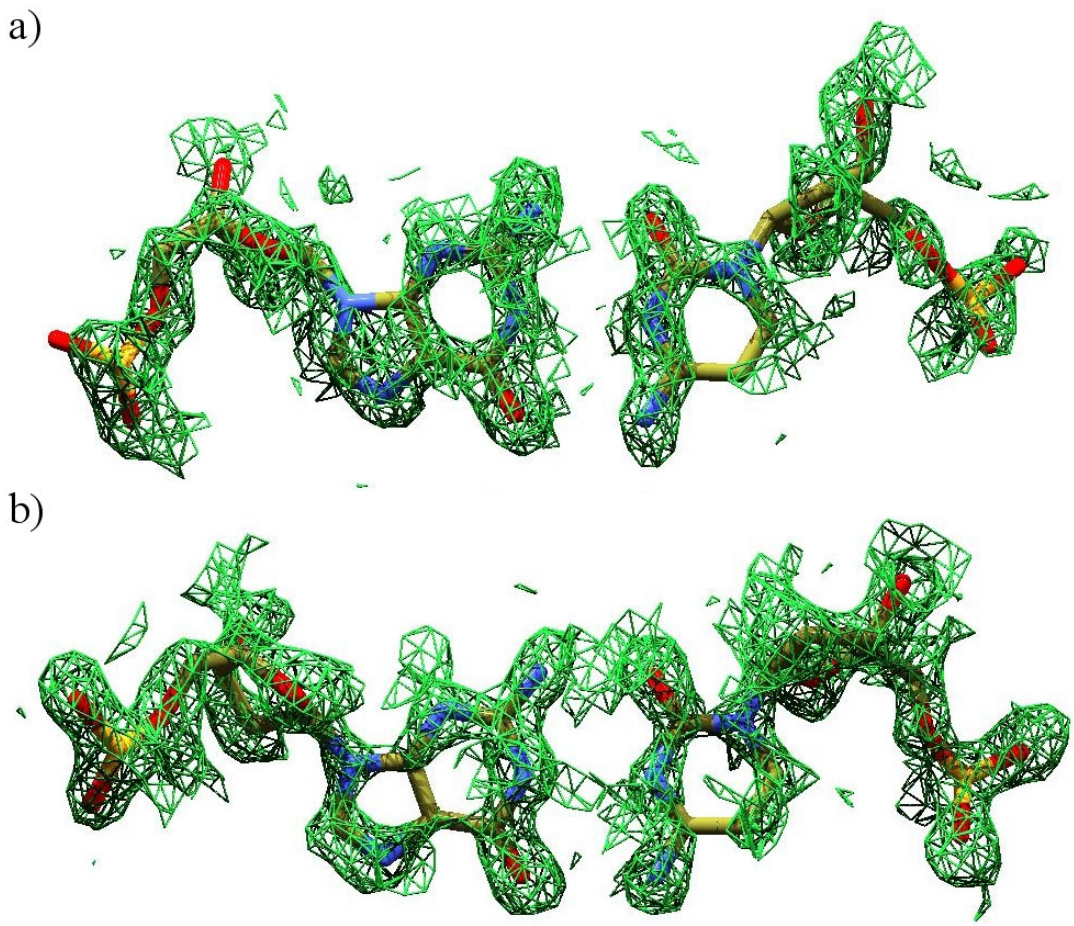

Figure 4. Sum electron density calculated from twinned data.
Poor sum electron density contoured to 1.0 σ is drawn in green net. The data were obtained by integration and scaling of all images obtained from the crystal structure of d[(CGCGAATXCGCG)]$_2$ (X = Z3dU). The figure shows the a)G4-C22 and b)C10-G15 basepairs.

values that are significantly different from the previous and next frames. Table 2 lists the average intensity, average error in intensity, χ^2, and R_{sym} values for these two data sets. Detailed data processing statistics generated by HKL2000 can be found in appendix I.

Calculation of Anomalous Density Maps

Locating heavy atom sites using their anomalous scattering is similar to a Single-wavelength Anomalous Diffraction (SAD) experiment. The goal of a SAD experiment is

Table 2. Summary of data processing statistics from HKL2000.

Resolution Range (Å)		Average Intensity	Average Error	Normal χ^2	R_{sym} [a]
[d(CGCGAATXCGCG)]$_2$					
50.00	3.21	30204.8	565.5	2.765	0.044
3.21	2.55	10244.6	189.2	1.903	0.046
2.55	2.23	3984.8	83.8	1.543	0.050
2.23	2.02	2754.9	66.0	1.335	0.053
2.02	1.88	1598.4	49.1	1.040	0.060
1.88	1.77	1119.2	42.3	1.029	0.081
1.77	1.68	856.7	39.5	0.884	0.086
1.68	1.61	610.4	36.6	0.739	0.110
1.61	1.54	286.8	34.8	0.697	0.187
1.54	1.49	202.3	58.9	0.639	0.238
All reflections		5359.7	118.9	1.289	0.050
[d(CGCGAAXXCGCG)]$_2$					
50.00	3.21	15844	281.4	2.165	0.041
3.21	2.55	3781.5	79.2	1.230	0.043
2.55	2.23	1089.9	34.0	0.891	0.058
2.23	2.02	513.1	27.4	0.689	0.086
2.02	1.88	262.2	25.0	0.594	0.144
1.88	1.77	172.5	27.3	0.529	0.200
1.77	1.68	181.7	33.3	0.525	0.171
1.68	1.61	163.9	36.8	0.511	0.150
1.61	1.54	89.6	31.5	0.474	0.272
1.54	1.49	36.0	41.5	0.445	0.558
All reflections		2976.4	71.2	0.959	0.047

[a] $R_{sym} = \frac{\sum_N \left[\sum_n \left(I_n(h) - \bar{I}(h)\right)/n\right]}{\sum_N \bar{I}(h)}$,

where N is a set of symmetry related reflections, $I_n(h)$ is the n^{th} intensity measurement of reflection h in this set, $\bar{I}(h)$ is the weighted-average of all intensity measurements for reflection h.

to use the positions of anomalous scatterers to calculate phase angles. In SAD, the unknown information is the phase angles of the scattered waves, while the known information is the location of the heavy atoms (obtained from the data). Here the problem is reversed. The known information is the phases of the scattered waves (calculated using a model) and the unknown is the location of heavy atoms.

Anomalous ($|F^+| - |F^-|$) electron density maps were calculated using two different methods (Figure 5). *Method A* employs the entire model for initial phase calculation while *method B* uses only the positions and the occupancies of the phosphorous atoms. It is essential to note that the fundamental theory for both methods is absolutely the same. Table 3 lists the anomalous scattering factors f'(λ) and f''(λ) at two wavelengths. Since the only atoms in the starting model with significant anomalous scattering are the phosphorous atoms, the two methods are essentially the same.

Although the data sets used in this thesis were collected at the two wavelengths listed in Table 3, the ideal wavelength is at the appropriate edge of the heavy atom in the crystal. Therefore, the collected data contain a weaker anomalous signal than data collected at the appropriate wavelength. However, even the weak anomalous signal was detectable and readily identified the position of the anomalous scatterer. It is possible that data collected at the thallium or rubidium edge would result in identification of additional sites for Rb^+ or Tl^+. The rubidium $f'(\lambda)$ and $f''(\lambda)$ values at its K-edge wavelength are –8.215 electrons and 3.770 electrons, respectively. The thallium $f'(\lambda)$ and $f''(\lambda)$ values at its L_{III}-edge are –16.578 electrons and 10.134 electrons, respectively. In addition, the phosphorous anomalous signal at these wavelengths is lower. Several trials with different data sets showed that *method B* produces anomalous electron density maps that are much more clear and easier to interpret. The most intense peaks in the anomalous electron density maps calculated by either method were associated with the phosphorous atoms. The intense phosphorous peaks are expected and show that (i) the two methods produce the same results (ii) the phosphorous peaks can be used as a test for validating the accuracy of the anomalous maps.

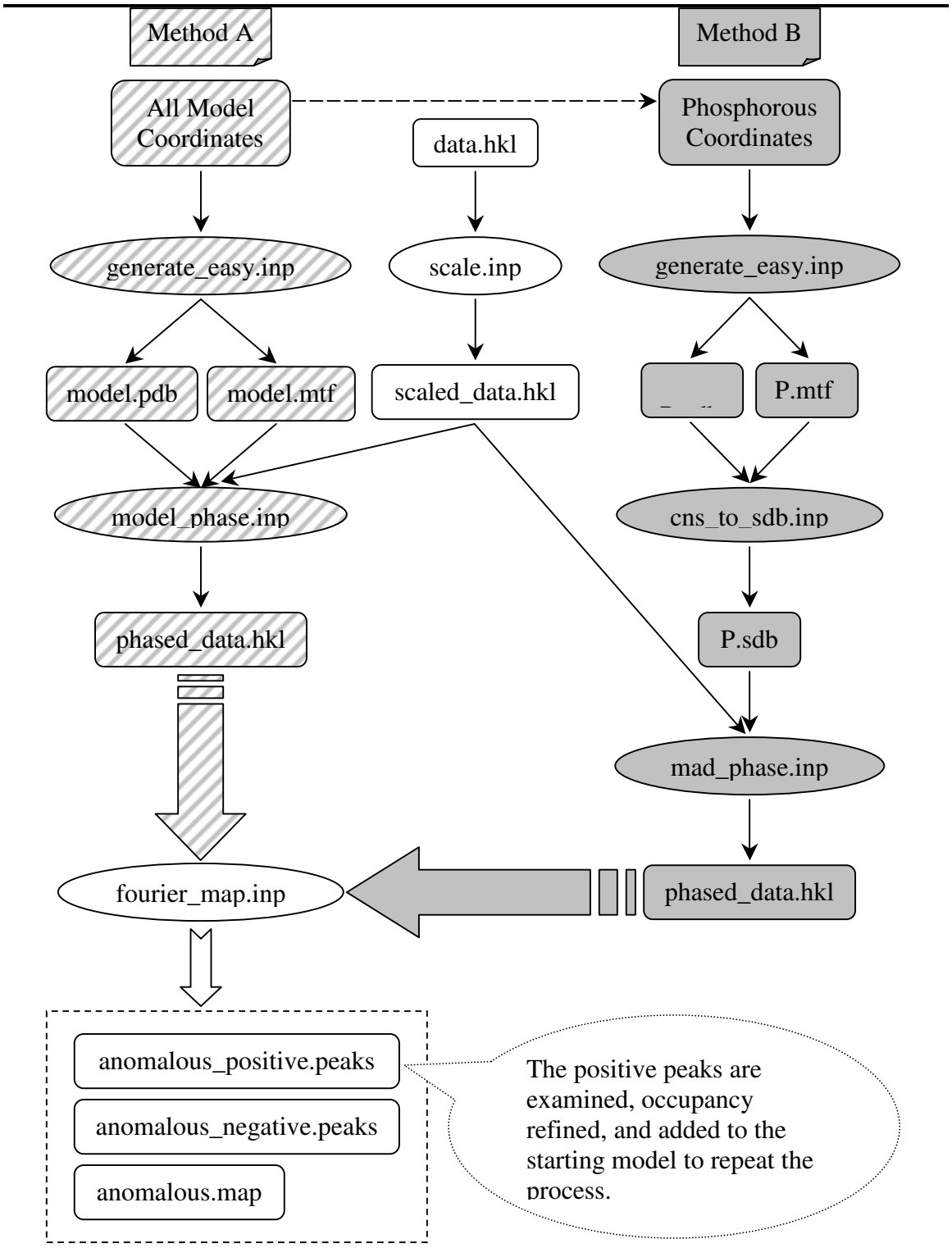

Figure 5. Two methods for calculation of anomalous density maps in CNS.
All CNS (Crystallography and NMR System) input script files are in ovals all text data files (coordinates and structure factors) are in squares. Method A uses the entire model and is shown in striped grey background. Method B uses the position of the phosphorous atoms and is shown in solid grey background. Data preparation and the final calculation of the density map is the same for both methods and is shown in white background.

Table 3. Anomalous scattering factors.

	$\lambda = 1.5405$[a] Å		$\lambda = 1.0090$ Å	
Element	$f'(\lambda)$ (e)[b]	$f''(\lambda)$ (e)[b]	$f'(\lambda)$ (e)[b]	$f''(\lambda)$ (e)[b]
C	0.017	0.009	0.007	0.004
N	0.029	0.018	0.012	0.007
O	0.046	0.032	0.020	0.013
Na	0.130	0.124	0.064	0.052
Mg	0.165	0.177	0.084	0.075
P	0.284	0.433	0.165	0.191
Cl	0.347	0.702	0.227	0.317
K	0.366	1.066	0.285	0.490
As	-1.010	1.005	-2.591	3.589
Rb	-0.574	1.608	-1.474	0.750
Cs	-1.021	7.902	-0.270	3.897
Tl	-4.881	8.085	-9.991	4.114

[a]CuKα radiation from in-house source. [b]f' and f'' are in units of electrons (e).

Several notes must be made about the common problems that arise in CNS (Crystallography and NMR System). The Wilson scaling performed by the script input file 'scale.inp, is very important and is in addition to scaling performed during image processing. It's generally useful to obtain some statistics about data before starting the process. The script file 'analyze.inp' can be used to obtain details statistics. Occasionally, the positive anomalous peaks selected by CNS in the final step are scattered throughout the structure and are difficult to interpret. In this case, the peaks can be shifted to a symmetry related position closer to the DNA molecule by applying the appropriate symmetry operator. CNS can be used to accomplish this task using the script 'shift_sites.inp'; however, the peak file must be converted to a PDB format file. The SDB file generated by 'cns_to_sdb.inp' script will have an empty 'seg_id' field, and may cause a problem in the next step. An arbitrary value can be assigned to this field.

Calculation of contacting distances between two sets of atoms was performed using a program written in Perl (Appendix II). The reason for this approach was an apparent error in the 'contacts.inp' script provided in the CNS program suit. The Perl program in appendix II does not generate the symmetry. Therefore, it was necessary to generate the symmetry equivalent models in the immediate neighborhood of the original model using the CNS script 'neighbors.inp'. The original model and the symmetry related models were merged into one file using the script 'merge_structure.inp'. The final coordinate file containing the original and all symmetry related models was used as the input coordinate file for the contact program in Appendix II.

CHAPTER III

AMBIGUITY OF IDENTITY AND OCCUPANCY: ANALYZING THE SOLVENT REGION OF DNA X-RAY CRYSTAL STRUCTURE

Abstract

This chapter is the first draft of a manuscript prepared for publication with the aid of Dr. Loren D. Williams. Here we explore the subjectivity intrinsic to macromolecular crystallography, focusing on the hydration/counter-ion region of nucleic acids. Water molecules, monovalent and divalent cations, and polyamines compete for similar or adjacent sites. Many of these species produce identical electron distributions (electron density maps), creating what we have termed an *identity ambiguity* in the solvent/counter-ion region of crystal structures. The *identity ambiguity* arises from the fact that scattering iso-types allow one to construct different models that produce similar fits of model to data. Even models with different electron densities can result in similar fits of model to data because various parameters compensate. The geometries of the coordinating ligands of many species (magnesium excluded) are similar to each other, and are effectively identical within the limitations of macromolecular crystallography. The observed distances are commonly occupancy-weighted averages. We test for the existence of this *identity ambiguity* by creating several models from previously published DNA structures with different scattering iso-types as the solvent species and comparing the resulting electron density maps and refinement statistics.

The occupancies assigned to the atoms and molecules in the solvent/counter-ion region are arbitrary to a certain degree. The refinement process adjusts various

parameters, compensating for over- and under-occupied regions. We term the arbitrary equality of the occupancies in the solvent/counter-ion region the *occupancy ambiguity* and perform a set of refinements to test the veracity of our conjecture.

In sum, atom-type assignments and occupancies in the hydration/counter region cannot be unambiguously extracted from conventional x-ray diffraction experiments. We present several examples where incorrect atom-type assignments have been revealed by anomalous scattering experiments. Published structures and the database that contains them do not provide realistic representations of objectivity. Reconsideration of macromolecular model-building protocols may be in order. Anomalous scattering provides information that can allow one to characterize the hydration/counter-ion region with greater accuracy than was previously possible.

Introduction

The subjectivity inherent in macromolecular crystallography is not always apparent to end-users, or even practitioners of crystallography. Here we wish to focus on the dark side, and to discuss and illustrate errors and uncertainty. The focus is the hydration/counter-ion milieu of nucleic acids, including DNA and RNA oligonucleotides, DNA-drug complexes, DNA-protein complexes, and ribosomes. For nucleic acids and their complexes the hydration/counter-ion region is especially non-homogeneous because the high anionic charge induces association of many types of cations. Water molecules, monovalent and divalent cations, and polyamines compete for binding. Therefore, many of the issues discussed here are relevant to nucleic acids, but not necessarily to proteins. The goal is to stimulate new approaches, and to promote cautious and reasoned

interpretation of structural information. We believe reconsideration of model-building, reporting and archiving protocols is in order.

The method scrutinized here is the assignment of water molecules and ions to isolated sum and difference electron density peaks ($2|F_o|-|F_c|$ and $|F_o|-|F_c|$ Fourier peaks) during the latter stages of refinement. The standard protocol is 'when in doubt, make it a water', with occupancy 1.00. If the thermal factor and geometry of the water molecule remain reasonable after refinement, then the water atom-type assignment and occupancy of 1.00 are considered to be validated. As explained below, advanced experimental approaches (anomalous scattering) demonstrate that these methods are not reliable. In the absence of anomalous scattering data, alternate atom-type assignments are possible, giving many models with reasonable thermal factors and geometry that fit the data equally well. The resulting models contain what we have termed the *identity ambiguity*. Furthermore, the occupancies assigned to the molecules and atoms in the solvent/counter-ion region are 'flexible'. Various parameters can be adjusted to compensate for hyper-occupancy (excess of electron density in the model compared to in the data) and hypo-occupancy (lack of electron density), creating an *occupancy ambiguity*. Selecting one model, with a given composition of the hydration/counter-ion region, over another model with a different composition is essentially a subjective enterprise, and is not grounded in the experiment.

To validate our hypothesis, we carried out a series of refinements on two previously published B-DNA structures. In one set of refinements, we empirically tested our conjecture of *identity ambiguity* in the solvent/counter-ion region of crystal structures of nucleic acids. All water molecules in these models were changed to fully occupied

Na^+ or 56% occupied K^+ ions and the models were refined against the data. In order to test the veracity of the *occupancy ambiguity*, the total occupancies of the water molecules in the two structures were systematically reduced and the models refined against the data. The results obtained overwhelmingly support the subjective nature of the hydration/counter-ion regions of both models. In the concluding section we give some examples from our own work illustrating how the conventional approach can go very wrong.

Materials and Methods

Substitution of water molecules with Na^+ and K^+ iso-type.

Coordinates and structure factors for the low Mg^{2+} forms of the Dickerson-Drew Dodecamer (DDD) grown in presence Na^+ (DDD-Na^+, entry BDL084) and K^+ (DDD-K+, entry BD0005) were obtained from the NDB (Shui et al., 1998a; Shui et al., 1998b). Each set of coordinates was used to construct two additional models. In one model the water molecules were substituted with Na^+ ions, while in the second model the water molecules were substituted with 56% occupied K^+ ions (Figure 6).

The occupancies of the K^+ and Na^+ ions were adjusted relative to the original occupancies of the water molecules. Hence a 50% occupied water molecule was substituted with a 50% occupied Na^+ ion in one model and 28% occupied K^+ in the other model. The result was two sets of models containing three structures with different molecular and atomic iso-types in the hydration/counterion region. All models were annealed and refined against their corresponding data sets using the program CNS

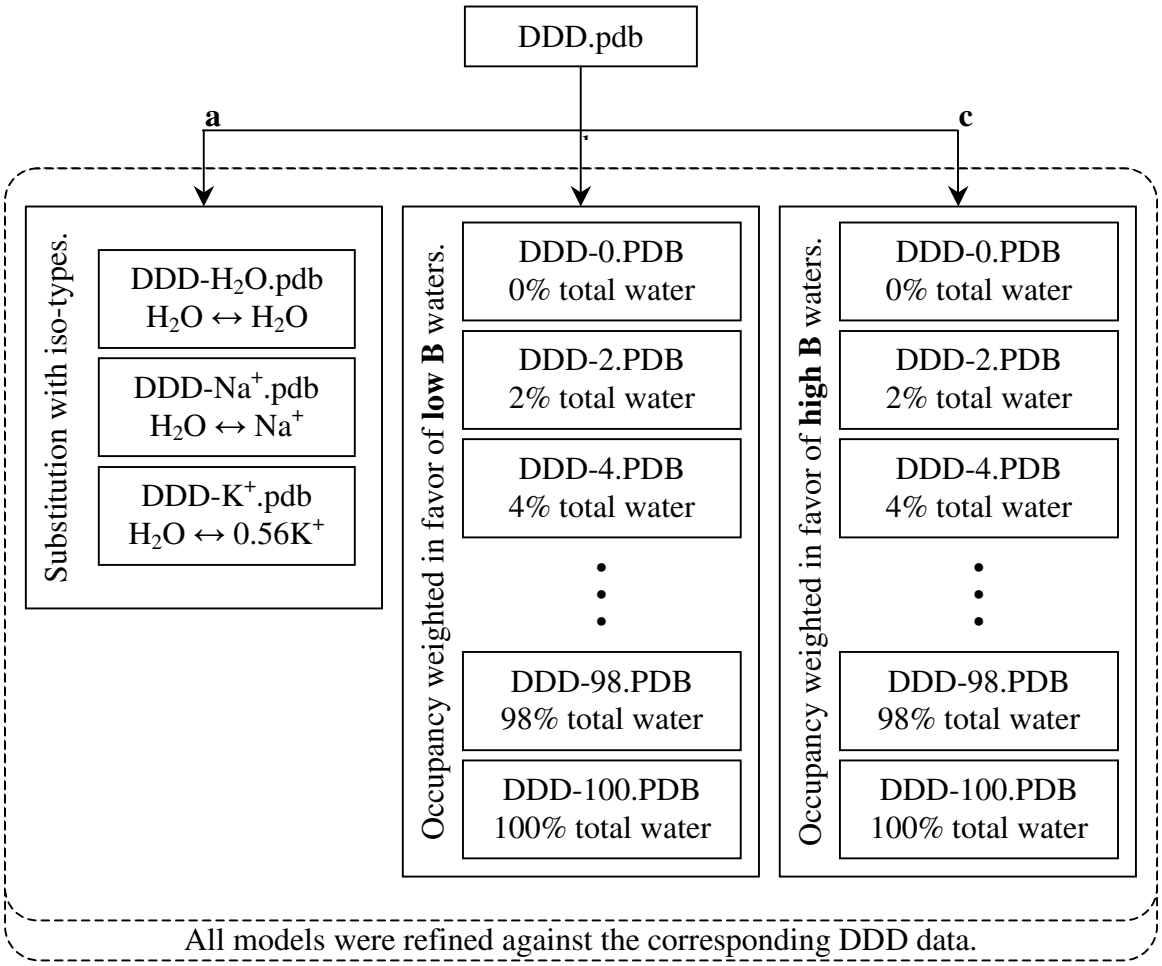

Figure 6. The experimental scheme for testing the identity and occupancy ambiguity. 'DDD.pdb' represents the coordinate file for either the DDD-Na^+ or the DDD-K^+ structures. The data used were the original data submitted with each structure. All models derived from DDD-Na^+ were refined against the DDD-Na^+ data set, and similarly, models derived from DDD-K^+ were refined against the DDD-K^+ data. Each starting model was used to create three models with water iso-types substitutions (a), and 51 models with step-wise reduction of total occupancy of water molecules using either W_{low} (b) or W_{high} (c).

version 1.1 and parameters of Berman and co-workers. Table 4 contains the relevant statistics.

Reduction of the total occupancy of water molecules.

To explore correlations between *occupancy ambiguity* and thermal factor, we created a weight scheme based on the thermal factor of the water molecules. For each of

the two starting structures above, two sets of models with 0% to 100% total water occupancy in increments of 2% were created. Each set contained 51 models. The occupancies of water molecules were distributed based on their individual temperature factors. One set was weighted using weight factor, W_{low} favoring water molecules with low temperature factor, while the second set used W_{high} to weight the occupancies towards waters with high temperature factor (Figure 6). The weight factors W_{low} and W_{high} are

$$W_{low} = \frac{B_i - B_{\min}}{\sum (B_i - B_{\min})}$$

$$W_{high} = \frac{B_{\max} - B_i}{\sum (B_{\max} - B_i)}$$

where W_i and B_i are the weight factor and temperature factor for the i^{th} water molecule, and $B_{\min}$ and $B_{\max}$ are the minimum and maximum observed temperature factors for all water molecules, respectively. Each set of models was refined against the corresponding data in the same manner as before.

Results

Testing the identity ambiguity.

Substitution of water molecules with Na^+ and K^+ iso-types did not lead to a significant increase in *R*-factor and *R*-free for any of the models derived from either DDD-Na^+ or DDD-K^+. The highest increase was seen from substitution of water with Na^+ in both structures, resulting in 2.3% and 3.5% increase in *R*-factor and *R*-free respectively (Table 4). Substitution of water with 56% occupied K^+ had virtually no effect at all. The same result was seen in the rms deviation in bond distances and angles.

Table 4. Refinement statistics for DDD models with water, Na^+, and K^+ iso-types.

	DDD-Na+			DDD-K+		
unit cell						
a (Å)	25.186			25.296		
b (Å)	40.208			40.244		
c (Å)	65.656			65.939		
space group	$P2_12_12_1$			$P2_12_12_1$		
DNA (asymmetric unit)	$[d(CGCGAATTCGCG)]_2$			$[d(CGCGAATTCGCG)]_2$		
major couterion present	Na^+			K^+		
no. reflections	58307			61696		
no. of unique reflections	11594			11376		
no. of reflections used	11416			6637		
max obsd. resolution (Å)	1.38			1.50		
max resolutions used (Å)	1.40			1.75		
no. of DNA atoms	486			486		
no. of solvent molecules[a]	142 (full), 18 (half)			126		
no. of magnesium ions plus 1st shell water[b]	7			7		
no. of spermine atoms	7			7		
	H_2O	Na^+	K^+	H_2O	Na^+	K^+
R-factor (%)	17.45	17.86	17.54	19.94	20.34	19.91
R-free (%)	19.48	20.17	19.75	25.94	26.67	26.21
rms deviation of bonds from ideality (Å)	0.0094	0.0091	0.0092	0.0096	0.0096	0.0095
rms deviation of angles from ideality (degrees)	1.4535	1.4269	1.4468	1.4902	1.4800	1.4868

[a]The number of original water molecules excluding magnesium first shell. [b]These water molecules were not converted to other iso-types.

For each set of three structures (100% H_2O, 100% Na^+ and 56% K^+), the rms deviation of positions between the structures was zero up to two significant figures. We compared the electron density maps in each set of three structures carefully and found no significant change. Special care was taken to look for negative and positive difference density changes and no difference was observed among the three models in each set. A representative case is shown in Figure 7. The sum and difference electron density maps are drawn around the solvent species in the 'spine of hydration' for the three models

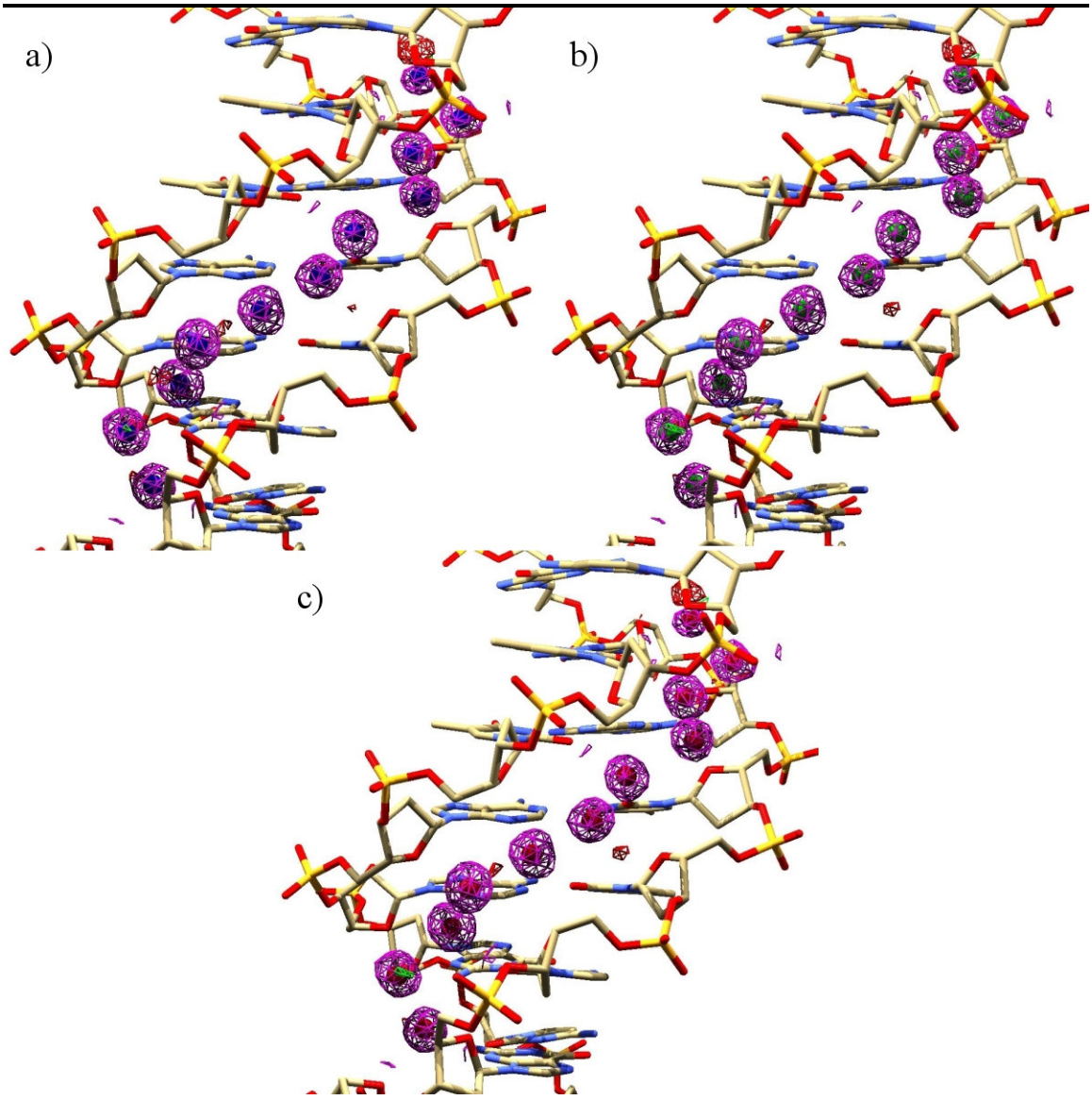

Figure 7. Electron density in three iso-types of the DDD-Na+ spine of hydration.
The sum (contoured at 1.0 σ, magenta net), positive difference (contoured at 2.5 σ, red net), and negative difference (contoured at –2.5 σ, green net) electron density are drawn around three iso-types of the DDD-Na+ structure's minor groove spine of hydration. Small amounts of difference density are scattered throughout each image. The spine is occupied with a) 100% Na^+, blue spheres b) 56% K^+, green spheres c) 100% water molecules, red spheres.

derived from DDD-Na^+. The electron density maps are unchanged when the water molecules are substituted with Na^+ or K^+. The positions of the Na^+ or K^+ ions are not significantly different from the position of the water molecules (rmsd of 0).

Testing the occupancy ambiguity.

For each structure, there exists an occupancy threshold, above which the *R*-factor and *R*-free change slightly and independent of the temperature factor of the water molecules. In other words, above this threshold an increase in the total occupancy of the water molecules does not reduce the *R*-factor or *R*-free significantly and there is no observed correlation between thermal factor and total water occupancy. Below this threshold, reducing the occupancies of waters with high thermal factors results in a sharper increase in *R*-factor and *R*-free than reducing the occupancies of waters with low thermal factor. A plot of the *R*-factor and *R*-free values as a function of occupancy is shown in Figure 8. The thresholds are indicated with arrows. The total water occupancy threshold (relative to the original models) is 80% for the DDD-Na+ structure, which has a resolution of 1.40 Å, and 60% for the DDD-K+ structure with a resolution of 1.75 Å. For both structures, changing the occupancy of the water molecules to 85% of their original value resulted in a 0.49% increase in R-factor.

Discussion

Estimates of errors.

The macromolecular crystallographic method provides less information on errors, and less information to assess the fit of model to data, than is the norm for physical-chemical or biophysical fitting processes. To help the reader appreciate the distinctiveness of macromolecular crystallography, here we compare it with isothermal titration calorimetry (ITC) and small molecule crystallography.

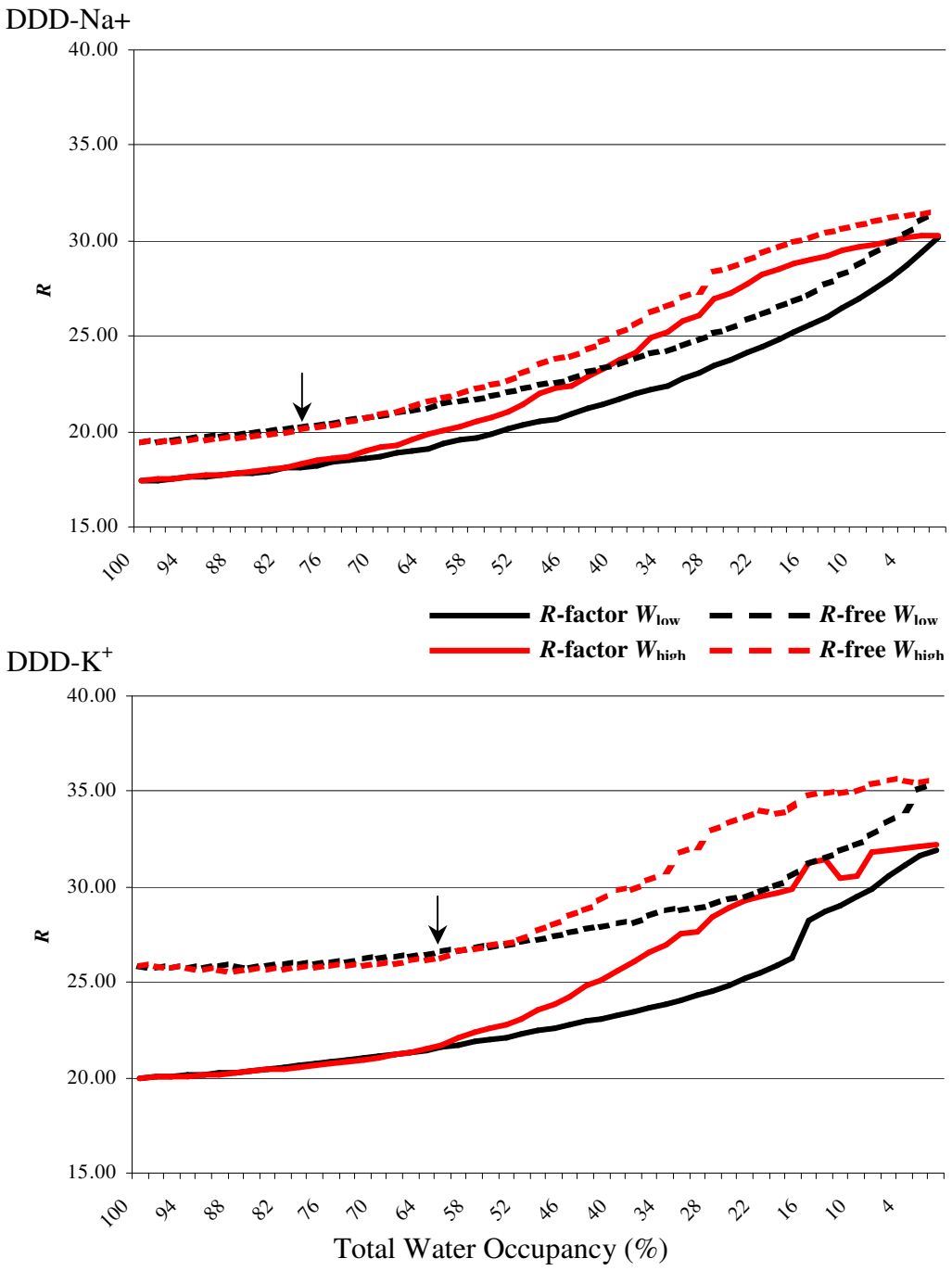

Figure 8. Plots of *R* vs. total water occupancy for DDD-Na+ and DDD-K+.
The *R*-factor (solid lines) and *R*-free (dashed lines) are plotted with respect to total water occupancy. The occupancies are weighed towards water molecules with low (black lines) and high (red lines) thermal factors using the weight factors W_{low} and W_{high}. The two arrow indicate the total occupancy thresholds.

Isothermal titration calorimetry.

In this method, one collects data (heats of injection) and conceives a model, with parameters such as an equilibrium constant, a stoichiometric coefficient and an enthalpy of binding. The parameters and the model are used to obtain calculated heats of injection. Differences between observed heats and the calculated heats are minimized by adjusting the parameters. During and after the fitting process one obtains a measure of global fit (χ^2), along with estimates of error of each parameter. The relationship between global fit and estimates of parameter error is not direct. The measure of global fit might be very good, even while estimate of error in one or more of the parameters is large, for example, if data on one side of the binding curve were absent.

Small molecule crystallography.

In this method, one collects data [observed structure factor amplitudes ($|F_o(hkl)|$)] and establishes a preliminary model. The parameters of the model are the x, y, z coordinates of various atom-types, their thermal factors, which are generally anisotropic, and their occupancies. The parameters are used to obtain calculated structure factor amplitudes ($|F_c(hkl)|$). Differences between $|F_o(hkl)|$ and $|F_c(hkl)|$ are minimized by adjusting the parameters, and often by adding or subtracting atoms from the model. As with ITC, correctness of the model is indicated by global measures of fit (the *R*-factor) in addition to estimates of error in individual parameters.

Macromolecular crystallography.

As in the small molecule method, in the macromolecular method differences between $|F_o(hkl)|$ and $|F_c(hkl)|$ are minimized by adjusting the parameters of the model. Both methods yield global measures of fit (*R*-factor, *R*-Free). However, realistic

estimates of uncertainties of individual parameters (the x, y, z coordinates of each atom, the thermal factors, and the occupancies) are not obtained from the macromolecular method. The origins of this deficiency are beyond the scope of this discussion, but are related to very large numbers of parameters and very large numbers of data. Macromolecular structures [entries in the Nucleic Acid Database (Berman et al., 1998)] do not specify uncertainties of individual x, y, z coordinates, their thermal factors or occupancies because those estimates of error are not available.

The protocol, after a reasonable model of the macromolecule along tight binding ligands is established, is to model the hydration/counter-ion region by adding water molecules and/or ions to corresponding sum and difference peaks of electron density (examples of sum electron peaks are shown in Figure 9). However, the electron density in the hydration/counterion region of nucleic acid crystals is intrinsically uninformative and ambiguous. Ambiguity arises (i) from mixed and partial occupancies, (ii) from scattering iso-types, and (iii) from parameter-compensation. Each of these effects is explained and discussed below.

Mixed and partial occupancies and scattering iso-types.

Crystallographic electron density and models are not restricted to formally correct chemical entities. The observed electron density is an ensemble average. Partially-occupied atoms and hybrid atom-types are 'observed'. This phenomenon is particularly acute in the hydration/counter-ion region of nucleic acids. Many species including water molecules, monovalent and divalent cations, and polyamines compete for similar or adjacent sites. It is possible for an electron density peak to arise from 40% one species, 30% another species, and 30% nothing.

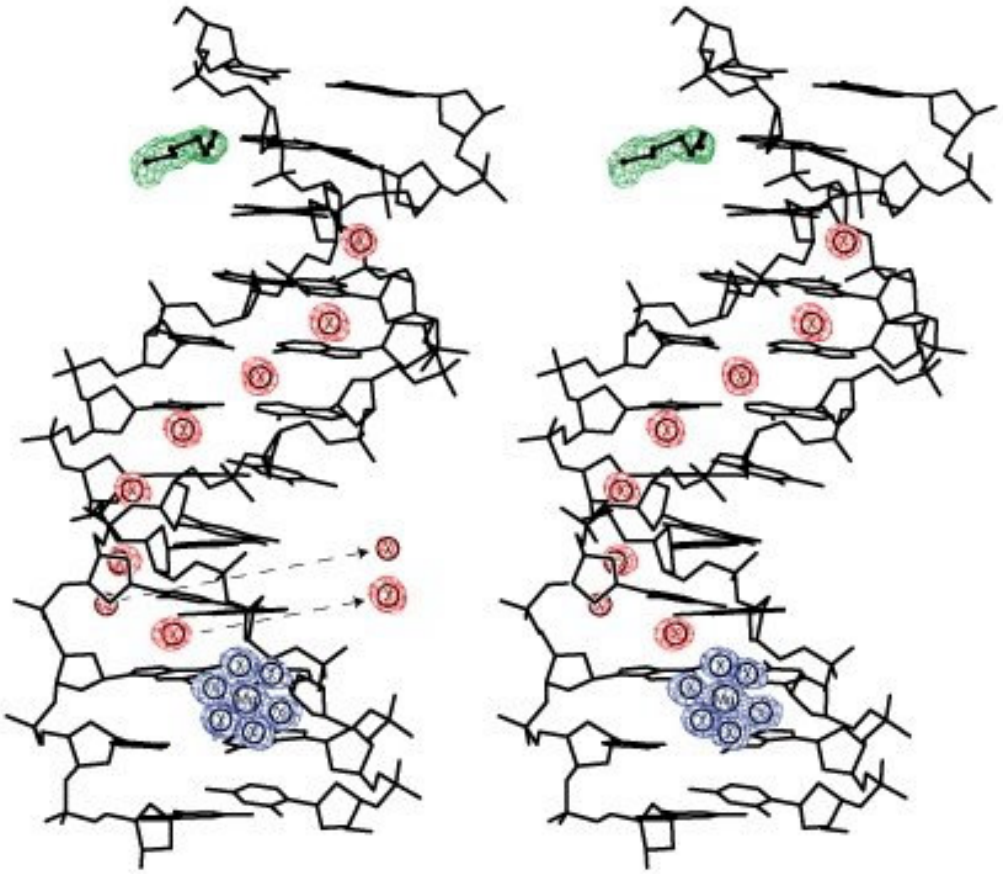

Figure 9. Stereoview of sum electron density from DDD-Na+ contoured at 1.0 σ
Only the electron density surrounding the hydrated magnesium ion (blue net), the partial spermine molecule (green net), and the minor groove solvent sites (red net) that interact with DNA bases is shown. For this structure, solvent sites were fitted to water molecules. The sum density from two minor groove solvent sites has been abstracted on the left side, to illustrate the lack of correspondence between the radius of a peak of electron density and the radius of the atom type.

A large number of chemically distinct models give identical electron distributions (electron densities). This degeneracy arises because many of the species in the hydration/counter-ion region are *scattering iso-types* (our term). Scattering iso-types are defined as chemically distinct species (molecules or ions) with indistinguishable electron distributions (and x-ray scattering properties). An ammonium ion and a water molecule are scattering iso-types. The number of electrons in each is the same (10 electrons). The

Table 5. Scattering iso-types: full and partial occupancies.

Species	Occupancy	Radius (Å)	Number of electrons
H_2O	1.00	1.40	10
Na^+	1.00	0.95	10
NH_4^+	1.00	1.40	10
Mg^{2+}	1.00	0.65	10
$K^+_{0.56}$ [a]	0.56	1.33	10
$Rb^+_{0.28}$ [b]	0.28	1.48	10
$Tl^+_{0.11}$ [c]	0.11	1.49	10
$Ca^+_{0.4}$ [d]	0.40	1.69	10

[a]60% occupied potassium ion. [b]28% occupied rubidium ion. [c]11% occupied thallium ion. d) 40% occupied calcium ion.

shape of the electron density is the same (spherical). A list of scattering iso-types is given in Table 5. Since partial and mixed occupancies are possible, the list of scattering iso-types is not restricted to formally correct chemical entities. The effective number of electrons of a potassium ion, when 55% occupied, is the same as that of a water molecule. In Table 5, occupancies that give equivalent numbers of electrons are indicated for common species. The list of possible scattering iso-types can be infinitely expanded by continuously varying the partial occupancies as illustrated for H_2O/Na^+ and H_2O/K^+ hybrids in Table 6.

We tested the sensitivity of the overall structure to the identity of the atoms and molecules in the solvent/counter-ion region by substituting all of the assigned water molecules in two structures with their scattering iso-types, fully occupied Na^+ and 56% occupied K^+. Of course, our substitutions are unrealistic; however, they serve to test the theories discussed here. The results in Table 4 show that the refinement process is insensitive to the identity of ion/molecules in the solvent region. The *R*-factor and *R*-free increased only marginally. There is minimal change in rms deviation of bond distances

Table 6. Scattering iso-types: mixed occupancies.

Mixed Occupancy H_2O/Na^+				Mixed Occupancy H_2O/K^+			
Occupancy H_2O	Occupancy Na^+	Effective radius (Å)	Effective number of electrons	Occupancy H_2O	Occupancy K^+	Effective radius (Å)	Effective number of electrons
1.0	0.0	1.40	10	1.0	0.00	1.40	10
0.9	0.1	1.40	10	0.9	0.05	1.36	10
0.8	0.2	1.39	10	0.8	0.10	1.31	10
0.7	0.3	1.39	10	0.7	0.15	1.27	10
0.6	0.4	1.38	10	0.6	0.20	1.22	10
0.5	0.5	1.37	10	0.5	0.30	1.18	10
0.4	0.6	1.37	10	0.4	0.35	1.13	10
0.3	0.7	1.36	10	0.3	0.40	1.09	10
0.2	0.8	1.35	10	0.2	0.45	1.04	10
0.1	0.9	1.34	10	0.1	0.50	1.00	10
0.0	1.0	1.33	10	0.0	0.55	0.95	10

and angles. The electron density maps showed no significant change. Figure 7 shows the spine of hydration for the three derivatives of the DDD-Na^+ structure. The greatest increase was in the individual temperature factors for the Na^+ and K^+ ions. Clearly, other parameters in the refinement are adjusted to fit the models with ions to the data.

One might naively assume that differences in shape, arising from differences in ionic/atomic radii could differentiate various species. The molecular radius of H_2O is greater than the ionic radius of Mg^{2+}, suggesting that the electron density from a Mg^{2+} ion should be of greater amplitude at the peak center and of less breadth than that of a water molecule. However, the effective dispersion of the electrons about the atom center, which is described in the structure factor equation as the rms amplitude of the atomic displacement, varies with thermal fluctuations, positional disorder, and data quality. The contributions of those effects generally obscures differences in ionic/atomic radii. This effect is illustrated in Figure 9.

Compensating parameters.

The problem extends beyond scattering iso-types and partial occupancies. Models with different electron densities can fit the same data equally well. Changes in one parameter are compensated in the fit by changes in other parameters. For example, decreasing the number of electrons of a K^+ ion (achieved by decreasing the occupancy or by changing the atom type from K^+ to Na^+ or H_2O) is compensated wholly or partially by decreasing the thermal factor. This compensation is related to determinacy (ratio of data to parameters). Therefore, hydration/counterion occupancies are generally not refined. Instead occupancies are fixed at a default value of 1.00 (the number of significant figures is absurd). This default value is arbitrary. In many cases switching an occupancy from 1.00 to 0.80 would not affect map quality or refinement statistics. Adjustments in thermal factors would eat up the difference.

The second set of experiments puts this idea to a test. Theoretically, it would be expected that the refinement statistics would be least sensitive to reduction in occupancy of the water molecules with high temperature factors. The data obtained supports this idea. However, in both structures a threshold in occupancy is observed, above which the increase in *R*-factor and *R*-free are independent of the temperature factors (Figure 8). A 20% reduction in occupancy of the water molecules resulted in a 0.7% increase in *R*-factor. To put this in perspective, there are several weight terms in the refinement process that can change the *R*-factor by as much as 1% at the expense of other refinement statistics. For example, in the program CNS the *R*-factor can be reduced by increasing w_a or decreasing r_{weight} (favoring the data over the restraints). Of course, the price of this reduction is a slight increase in rms deviation from ideality of bond distances and angles,

which is acceptable within limit since the restraints are for an ideal model. In fact, reducing the occupancies of the water molecules also results in a reduction of rms deviation for bonds and angles (data not shown).

Coordination Fingerprints.

As explained above the hydration/counter-ion region of nucleic acids is especially difficult to characterize unambiguously. Yet there is some hope provided by coordination fingerprints. The octahedral geometry and the ligand-to-metal distances of a well-ordered Mg^{2+} do allow quasi-reliable identification. It is conceivable that Na^{+} could be distinguished from H_2O by a coordination fingerprint. The ideal distance from Na^{+} to ligand is less than that from H_2O to ligand [Brown (Brown, 1988) has provided useful surveys of coordination geometries]. However, in the macromolecular experiment, the experimentalist must interpret subtle differences in geometries because the observed distances are generally occupancy-weighted averages (Table 6). How does one interpret nominally short contacts between water molecules? Do they arise from partial Na^{+} occupancy or simply from coordinate error? The database contains very few examples of Na^{+} ions with unambiguous coordination geometry, many with unreasonable geometry, and none with reasonable estimates of atom-type error. Many if not most of the Na^{+} ions in the NDB have coordination geometries consistent with those of water molecules.

Revaluation of published structures.

Considering the limitations imposed by mixed and partial occupancies, scattering iso-types, compensating parameters, and occupancy-weighted geometries it is simply impossible to extract unambiguously the correct identities of the contributing species

from the data (in the absence of anomalous scattering data). The hydration/counter-ion region cannot be reliably characterized by standard macromolecular x-ray methods. We believe there is a realistic possibility that many or even most 'water molecules' and ions are mis-assigned and/or incorrectly described in the NDB. In the following sections we provide some examples where the mis-assignment of atom-types has been experimentally revealed.

Mis-assigned peaks: Z-DNA.

A very high resolution structure (1.0 Å resolution) of the spermine form of Z-DNA (Egli et al., 1991) was determined by Martin Egli, Loren Williams and Qi Gao, as postdocs in the laboratory of Alexander Rich. Water molecules were assigned to sum and difference electron density peaks, including a series of regularly spaced peaks in the minor groove, which formed a 'spine of hydration'. The water molecules were well-behaved, with reasonable geometry and interactions, realistic thermal factors, good electron density, etc. The final published structure (NDB entry ZDF029) contains water molecules within the 'spine of hydration', which are assigned atom types of O (hydrogen atoms are implicit).

After that structure had been completed and published, evolution of cryo-crystallographic techniques allowed the same group to collect the first ever data set from a flash-frozen crystal of DNA (Bancroft et al., 1994), consisting of the same spermine-form Z-DNA. Lowering the temperature dampens molecular motions, increasing the information content of the diffraction data. The data obtained from a flash-frozen crystal indicated that minor groove contained a spermine molecule. Lowering the temperature converted the isolated peaks in the minor groove that had been assigned to a 'spine of

hydration' into a tube of electron density. The 'spine of hydration' morphed into a spermine molecule. At room temperature, the methylene groups of the spermine molecule were thermally disordered, and so not visible in the electron density maps. The amino groups of the spermine formed hydrogen bonds to the floor of the minor groove and were ordered at room temperature, and appeared as spheres of electron density. The peaks of electron density in the minor groove had been mis-assigned during the original refinement of the room temperature structure. They were not water molecules but were the amino groups of a spermine molecule. Our retrospective assessment is that our original conclusion, that peaks of electron density indicated a 'spine of hydration', was an over-interpretation of the data. However, even the revised structure, with occupancies of 1.00 for the spermine molecule, etc., is also an over-interpretation.

Mis-assigned peaks: B-DNA #1.

By 1998 there were sixty eight isomorphous members of the CGCXAATTYGCG (X = G or A, Y = C or T) dodecamer family in the NDB. Collectively, those structures contain thousands of water molecules, and no monovalent cations, divalent cations, or polyamines (Dickerson). We proposed an alternative interpretation of the data (McFail-Isom, Sines & Williams, 1999; Shui et al., 1998a; Shui et al., 1998b) in which the grooves of the B-DNA in those crystals are decorated with various types of cations. That proposal is now accepted, based in part on substitution experiments with scattering cations that give distinctive (anomalous) scattering information (Howerton et al., 2001; Tereshko et al., 1999; Woods et al., 2000) [also see work by Egli and coworkers (Tereshko et al., 2001)].

It is now clear that much of the hydration/counterion region contains partial and mixed-occupancy water molecules and cations. The results of anomalous experiments indicate that the A-tract minor groove contains monovalent cations. Increasing data quality has revealed that the major groove is associated with hydrated magnesium ions and spermine molecules (Figure 7). Our retrospective assessment is that the original models, which for example assumed that peaks of electron density indicated a well-ordered 'spine of hydration' (Drew & Dickerson, 1981; Kopka et al., 1983), were over-interpretations of the data. The predominant monovalent cation in the crystallization solutions was Na^+ (a scattering iso-type with water). Models with Na^+/H_2O partial occupancy hybrids would therefore fit the data equally as well as models with water only. There was never any experimental basis for setting the atom-types as pure water or their occupancies as 1.00.

Mis-assigned peaks: B-DNA #2.

In 1998 we identified a Mg^{2+} ion in the major groove of DDD (Shui et al., 1998a). The ion was well-behaved in the refinement, with excellent electron density, thermal factors, geometry, etc. The location of the Mg^{2+} was confirmed by other investigators (Chiu, Kaczor-Grzeskowiak & Dickerson, 1999; Minasov et al., 1999). It is a 'consensus' cation, observed in structures obtained from a variety of crystallization conditions and DNA modification. The major groove Mg^{2+} has been assigned a structural role by several investigators, and is thought to contribute to the famous 'dodecamer bend'.

In 2001 we discovered with an anomalous scattering experiment that the major groove Mg^{2+} is not fully occupied (Howerton et al., 2001). A monovalent signal is subtle

in the sum and difference maps but is unambiguous in the anomalous map. The monovalent site is displaced somewhat from the divalent site, although the proximity is such that occupancy of either site by a cation would preclude occupancy of the other (the sum of the occupancies cannot exceed 1.00). Since the Mg^{+2} is only partially occupied, the role of the Mg^{+2} in contributing to the dodecamer bend is unclear. Our retrospective assessment is that our original major groove Mg^{2+} ion model was an over-interpretation of the data. The good thermal factor/good geometry criterion was not sufficient to ascribe an occupancy of 1.00 to the Mg^{2+}.

Mis-assigned peaks: DNA-drug Complexes.

In x-ray structures of several early DNA-anthracycline complexes, well-defined and fully occupied Na^+ ions mediate interactions between the intercalated chromophore and the DNA (Frederick et al., 1990; Wang et al., 1987). In considering the reliability of those Na^+ assignments the following factors require consideration: (i) The Na^+ atom-type assignments are based on geometric considerations. It was assumed that six ligands surrounding a electron density peak is a definitively indicator of Na^+ ion occupancy. (ii) After initial refinement, the geometry of the ligands surrounding the Na^+ peak was restrained, to 'ideal' octahedral Na^+ geometry. (iii) The definition of 'ideal' Na^+ geometry has evolved over time. Six-coordinate octahedral geometry is no longer considered a reliable indicator of sodium. As noted by Jeffery (Jeffrey, 1997), water molecules engaging in bifurcated hydrogen bonding can be six-coordinate. (iv) Structures of some DNA-anthracycline complexes lack localized cations (Lipscomb et al., 1994; Williams et al., 1990). The Na^+ ions in the original structure have not proven to be fully reproducible. (v) Recent anomalous experiments shown that there are

additional monovalent cation sites in the DNA-anthracycline complexes (Howerton et al., 2003). In the anomalous experiments the original Na^+ sites are not as highly occupied as other sites.

Our retrospective assessment that localized cations and electrostatic forces are indeed important in structure, thermodynamics, and sequence specificity of DNA-ligand complexes. Favorable interactions of adriamycin and cations with the sequence-specific electrostatic landscape of DNA may be universal characteristics of DNA-small molecule interactions and may be useful in sequence-specific ligand design. However, the certitude and the details of the original Na^+ descriptions are in error.

Summary.

Water molecules, monovalent and divalent cations, and polyamines compete for similar or adjacent sites in the hydration/counterion regions of nucleic acids. In general, crystallographic models of hydration/counterion regions are biased simplifications, taken to a level of near criminality. In assessing the reliability of crystallographic models one must consider many factors. (i) With scattering iso-types, one can construct many different models with indistinguishable electron density maps. (ii) Even models with different electron density maps can give similar fits of model to data. (iii) The geometries of the coordinating ligands of common species (Mg^{2+} excluded) in the hydration/counter-ion region of DNA are very similar to each other, and are effectively identical within the limitations of macromolecular crystallography. (iv) The good thermal factor/good geometry criterion is not sufficient to ascribe an occupancy of 1.00. (v) Published structures and the database that contains them do not provide realistic representations of uncertainty. (vi) The atom type assignments and occupancies in the

hydration/counter region are generally subjective. (vii) In many cases electron density peaks should be assigned to wild-card atoms (10 electrons, identity unknown). (viii) Anomalous scattering experiments provide important information that allows one to characterize the hydration/counterion region with greater accuracy than was previously possible.

CHAPTER IV

COMPARISON OF THALLIUM (I) AND RUBIDIUM(I) AS ANALOGUES FOR BIOLOGICAL CATIONS IN STRUCTURES OF B-DNA

Abstract

The structure of DDD crystallized in the presence of Rb^+ (DDD-Rb^+) is solved to a resolution of 1.5 Å, and the anomalous scattering of rubidium is used to locate its positions in the structure. The anomalous scattering of the phosphorous atoms and their known coordinates in the structure are used for initial phase calculation. This structure is compared to structures of the same dodecamer grown in the presence of Tl^+ to compare and contrast the binding interaction of these cations with DDD.

We have also reviewed the chemical properties of K^+, Na^+, Rb^+, Cs^+, and Tl^+ comparing their possible interactions with B-form DNA. The specific valences of these cations are calculated for the ApT-step (P3 site) of DDD and compared to the best available estimates of their occupancy at this site. The estimates for the occupancies are derived from published x-ray and NMR structural data.

These results show that Tl^+ and Rb^+ are similar to each other and to other group I metals and in most applications can behave as a mimic for biologically relevant monovalent cations. However, there are subtle differences between Tl^+ and the group I metals that must be considered. This chapter is based on the first draft of a manuscript prepared for publication in collaboration with Dr Shelley Howerton and Dr. Loren D. Williams.

Introduction

DNA counterions along with other components of the 'hydration' milieu, are dynamic, and transit between a variety of states. Counterions with relatively small hydration enthalpies (Na^+, K^+, Rb^+, and Tl^+) interact directly with base and backbone functional groups. Counterions with larger hydration enthalpies, such as Li^+, and Mg^{2+} generally remain hydrated, and interact indirectly with DNA, via first-shell water molecules. DNA counterion interaction have been discussed in recent reviews (Hud & Polak, 2001; McFail-Isom et al., 1999).

It now appears that DNA interactions, both direct and indirect, influence positions and lifetimes of cations. Cation-DNA interactions cause differential localization of counterions in various 'territories' in the hydration region of DNA. Differential localization is supported by x-ray diffraction (Howerton et al., 2001; Shui et al., 1998a; Sines et al., 2000; Tereshko et al., 1999; Woods et al., 2000), NMR (Cesare Marincola et al., 2004; Hud & Feigon, 1997; Hud, Schultze & Feigon, 1998; Stefl et al., 2004), molecular dynamics simulations (McConnell & Beveridge, 2000; Young et al., 1997) and other techniques (Buckin et al., 1994). Cation distributions depend on DNA sequence and conformation. Cations localize preferentially in and around the minor groove of A-tracts and the major groove of G-tracts. Cation redistribution may be associated with conformational change, as for example a net migration from minor to major groove upon the B-form to A-form transition (Hud & Polak, 2001).

Experimental data to support or confute such models are highly dependent on the power of analytical methods to accurately characterize cation locations in the vicinity of DNA. X-ray diffraction retains serious deficiencies when used for the analysis of

solvent/counterion interaction even as it has evolved into an unparalleled method for characterizing macromolecular structure. The electron density in the hydration/counterion region of nucleic acid crystals is uninformative and ambiguous. Selecting one model, with a given composition of the hydration/counterion region, over another model with a different composition is a subjective enterprise, and is not grounded in the experiment [also, see the discussions by Williams (McFail-Isom et al., 1999; Shui et al., 1998b; Sines et al., 2000) and Egli (Tereshko et al., 2001)]. Ambiguity arises from mixed and partial occupancies, from scattering iso-types (H_2O, Na^+, NH_4^+, Mg^{2+}, partially occupied K^+, etc., produce indistinguishable x-ray scattering), and from parameter-compensation (thermal factor/occupancy) during model fitting. In some instances, cations can be differentiated from water molecules through the geometry of the first shell ligands; this approach is reliable primarily for divalent cations. In very few cases, monovalent cations have been identified in x-ray crystal structures, even then the coordination geometry suggests ambiguity.

The best method for identifying monovalent cations appears to be substitution with heavier cations. Historically Rb^+, Cs^+ and Tl^+ have been used as substitutes for Na^+ and K^+. (Badger et al., 1994a; Badger et al., 1994b; Bartenev et al., 1983; Basu et al., 1998; Gill & Eisenberg, 2001; Gursky et al., 1992a; Gursky et al., 1992b; Howerton et al., 2003; Howerton et al., 2001; Pedersen et al., 1998; Stroud et al., 1974; Tereshko et al., 1999; Tereshko et al., 2001; Villeret et al., 1995; Woods et al., 2000). If the chemical properties of the heavy monovalent cations are sufficiently similar to the light, biologically relevant cations, then they localize in similar sites with similar occupancies. Here we test the limits of the assumption of isomorphism by assaying several different

substitutes under conditions of crystallization, data collection, and refinement that are identical, as close as possible. A consensus among various substitutes would suggest the absence of significant differences among the substitutes, and therefore the absence of significant differences between the substitutes and the true target cation. Differences among substitutes would suggest differences between the substitutes and the true target cation. The goal here is to characterize and understand the differences between several heavy cation substitutes, by a comparison of structures that differ only in the identity of the substituted cation.

To perform the comparison here we control as many factors as possible in the crystal growth conditions, and data collection, reduction, refinement and analysis methodology. Previously anomalous diffraction has been used to characterize complexes of Rb^+ and Cs^+ with A-form DNA (Tereshko et al., 2001), and of Tl^+ with B-form DNA (Howerton et al., 2001) and DNA-drug complexes (Howerton et al., 2003). By contrast the non-anomalous difference Fourier methods were used to characterize complexes of Rb^+ and Cs^+ with B-form DNA (Tereshko et al., 1999; Woods et al., 2000). We hope to resolve whether some of the observed differences in occupancies and positions arise from different methodologies or from differences between the substitutes. For example, a single Rb^+ site was observed adjacent to dCGCGAATTCGCG, while 12 Tl^+ sites were observed.

The anomalous signal is of great utility for locating heavy atom including heavy ions. An anomalous map will show peaks of density at the positions of phosphorous atoms and of the heavy ion, but not at positions of the lighter atoms comprising the DNA or the solvent region. For a given atom, the strength of the signal depends on the

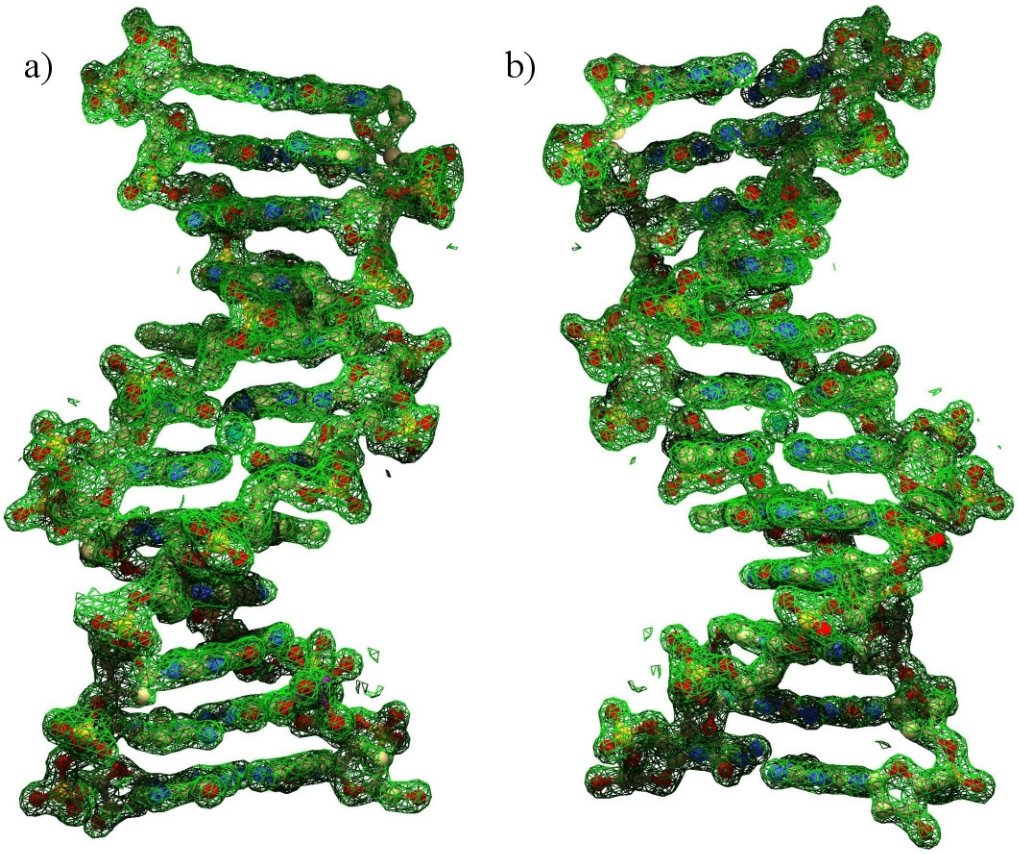

Figure 10. Sum electron density contoured at 1.0 σ around DDD-Rb^+.
The sum electron density contoured at 1.0 σ drawn in green net. The figure shows the view of minor (a) and major (b) grooves.

anomalous scattering factors [f'(λ) and f"(λ)], the occupancy, and the degree order. We compare the X-ray crystal structures of DDD obtained from crystals grown in the presence of Rb^+ (DDD-Rb^+) and from crystals grown in the presence of Tl^+ (DDD-Tl^+). The goal is to identify differences in location and occupancy between the binding of these ions to B-form DNA. Figure 10 shows the DDD-Rb^+ structure with sum electron density contoured at 1.0 σ.

Materials and Methods

Crystallization.

DDD-Rb^+ crystals were grown using the sitting drops vapor diffusion method. The crystallization solution contained 1.0 mM of ammonium salt of d(CGCGAATTCGCG) (Midland Certified Reagent Co., Midland, TX) purified by reverse-phase HPLC, 9.5 mM rubidium(I) cacodylate (pH 6.5), 10.4 mM magnesium acetate, 8.9 mM spermine acetate, and 3.8% 2-methyl-2,4-pentanediol (MPD). The crystallization solution was equilibrated against a reservoir of 50% MPD at room temperature. A crystal of dimensions 0.4 x 0.4 x 1.7 mm^3 grew within a week. Other crystallization experiments with higher concentration of Rb^+ produced either showers of small crystals or larger crystals that did not diffract well.

Data collection and processing.

The data were collected using an in-house Rigaku/MSC rotating anode generator (model RU-H3R) with Osmic blue confocal mirrors and a Raxis IV^{++} image plate detector. The x-ray radiation was a wavelength of 1.54 Å (Cu K_α). 360° (in ϕ) of data were collected, while the crystal was maintained at -180°C. MOSFLM 6.2.2 (Leslie, 1992; Powell, 1999) was used to index and integrate a total of 144823 data. SCALA (1994) was used to reduce and scale the reflections to a set of 20147 unique reflections keeping the Bijvoet pairs separate.

Refinement.

CNS version1.1 (Brunger et al., 1998) was used to merge the Bijvoet pairs to create a dataset containing 10905 reflections. DNA coordinates from a high-resolution

structure of DDD (NDB entry BDL084) (Shui et al., 1998a) were used as a starting model for molecular replacement. The model was annealed frequently throughout the refinement. Parameters of Berman and co-workers (Clowney et al., 1996; Gelbin et al., 1996; Parkinson et al., 1996) were modified by removal of dihedral restraints on the sugar-phosphate backbone. Careful inspection of electron density maps indicates that for high-resolution data, removing the backbone dihedrals facilitates correct sugar conformation. After each cycle of refinements, water molecules were added by selecting the strongest peaks in the sum and difference electron density maps, with the appropriate hydrogen bonding geometry.

Anomalous ($|F^+|-|F^-|$) density maps were calculated from the data containing unmerged Bijvoet pairs using CNS. The positions of the phosphorous atoms were used in conjunction with the f' and f'' values for phosphorous at 1.54 Å (Table 3, Chapter II) to calculate an initial anomalous map. Candidate sites for Rb^+ were identified from this map. Rb^+ atoms were added to the model and validated by refinement against the data and by geometric criteria. The coordinates for these sites were then used along with f' and f'' values for Rb^+ at 1.54 Å (Table 3, Chapter II) to generate a second anomalous map and search for additional Rb^+ sites. The occupancies of the Rb^+ sites were adjusted through successive refinements, until no positive or negative difference electron density was observed. Table 7 contains a summary of crystallization data and refinement statistics. The program Curves 5.3 was used to calculate helical parameters (Stofer & Lavery, 1994).

Table 7. Crystallization and refinement statistics for DDD-Rb^+ and DDD-Tl^+

	DDD-Rb^+	DDD-Tl^+
concentration of crystallization mixture		
DNA	1.0 mM	1.0 mM
spermine acetate (pH 6.4)	8.9 mM	8.9 mM
magnesium acetate	10.4 mM	5.2 mM
heavy atom	9.5 mM (Rb^+ cacodylate)	19.0 mM (Tl^+ acetate)
unit cell		
a (Å)	25.44	25.94
b (Å)	40.08	40.74
c (Å)	66.63	66.20
space group	$P2_12_12_1$	$P2_12_12_1$
temp. of data collection (°C)	-180	-110
no. of reflections	144823	131329(290894[a])
no. of unique reflections	10905	21760
completeness (%)/highest shell (%)	91/62	93/68
max resolution of obsd reflections (Å)	1.48	1.14
max resolution of highest shell	1.50	1.55(1.20[a])
used in refinement (Å)		
resolution range (Å)	35-1.5	35-1.55(35-1.2[a])
no. of reflections used in refinement	9312	18897
no. of reflections used in test set	1073	2100
rmsd of bonds (Å)	0.010	0.010
rmsd of angles (deg)	1.513	0.030
DNA (asymmetric unit)	$[d(CGCGAATTCGCG)]_2$	$[d(CGCGAATTCGCG)]_2$
no. DNA atoms	486	486
no. water molecules, excluding		
Mg first shell	135 full, 1 partial	116 full, 18 partial
no. Mg ions plus coordinating		
water molecules	7 full	7 partial
no. heavy cations/highest occupancy	1/0.52	13/0.35
no. spermine atoms	0	0
R-free (%)	24.1	22.2
R-factor (%) excluding test set	19.9	16.9

a) The anomalous data sets for DDD-Tl^+ and DDD-Rb^+ were collected using the same generator and detector . For DDD-Tl^+, a non-anomalous data set was collected to a higher resolution. The resolution and the number of reflections for this data set are reported in parenthesis. For both structures, the data were processed using the same software.

Results

The conformation of DDD-Rb^+ is in good agreement with an ensemble of DDD structures with a variety of different monovalent cations. This ensemble has previously been characterized, and is called the 'low [Mg^{2+}] DDD ensemble' by Sines et al (Sines et al., 2000). Analysis of the DDD-Rb^+ structure using the program Curves version 5.3 reveals that inter-base pair parameters of DDD-Rb^+, including helical rise, base pair roll helical twist closely follow the same trends as the other members of the low [Mg^{2+}] DDD ensemble (Figure 11).

The initial anomalous electron density of DDD-Rb^+ showed peaks for the 22 phosphorous atoms of the DNA, along with three candidate Rb^+ peaks in the solvent region. The anomalous electron density at one Rb^+ candidate site (the P3 site, at the ApT-step in the minor groove) was 3.84 σ, and was directly on top of a previously assigned water molecule. The other two anomalous density peaks were 3.69 and 3.46 σ, while the intensity of the phosphorous anomalous peaks ranged from 7.71 to 14.55 σ. Both of these anomalous peaks were greater than 1 Å from the nearest water molecule.

The model was modified by adding partially occupied Rb^+ ions to each of the three sites. This modified model was refined against the data followed by phase and map calculation. The Rb^+ ion located at the P3 site maintained reasonable geometry and remained within a sum electron density peak (Figure 12). The other two Rb^+ atoms were not observed to be within sum electron density, and did not maintain reasonable geometry. Therefore the Rb^+ at the P3 site remained in the model while the other two Rb^+ atoms were removed.

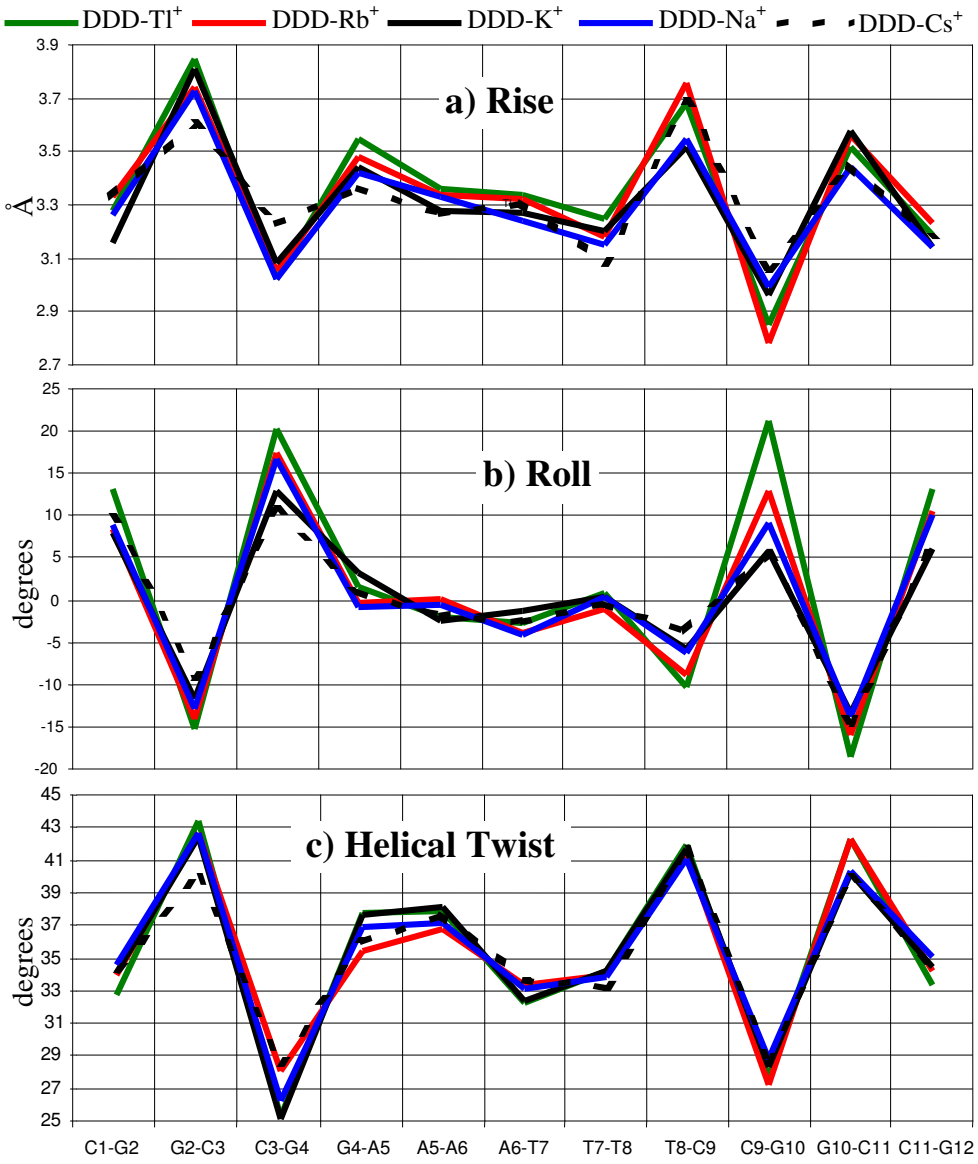

Figure 11. Plots of inter-basepair parameters for a selection of DDD structures.
Plots of inter-base pair rise (a), roll (b), and helical twist (c). All graphs have the same x-axis, marked at the bottem with the one set of bases in the corresponding basepairs. For example, C1-G2 is the parameters measured from C1G24 basepair to G2C23 basepair.

The final occupancy of the Rb^+ atom at the P3 site is 0.52. The occupancy was not determined by refinement, but was set empirically, by iterative inspection of sum and position and negative difference maps. Except at the P3 site the anomalous Rb^+ maps did not reveal evidence for peaks in the vicinity of the positions of Tl^+ atoms determined

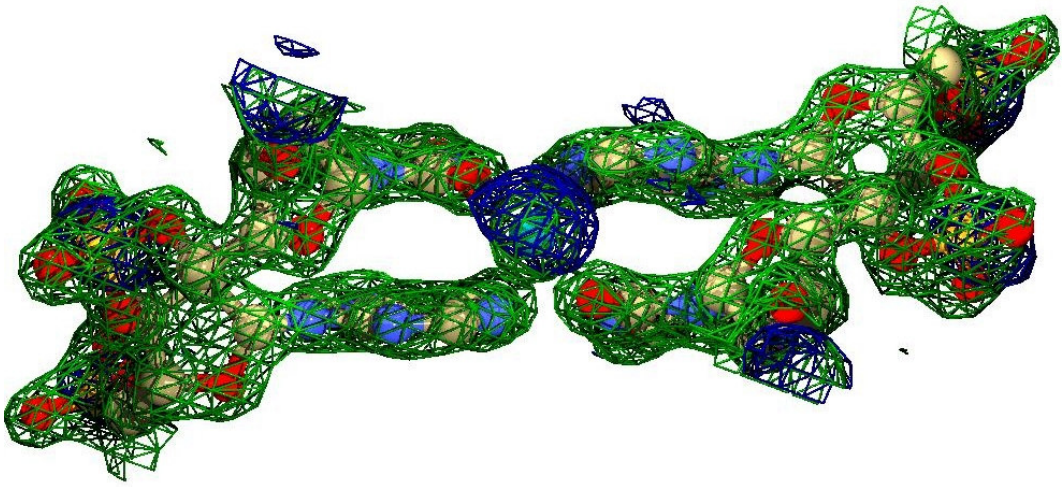

Figure 12. The sum and anomalous density contoured around the Rb^+ at the P3 site. Close-up view of the Rb^+ ion at the P3 site. The sum electron density contoured at 1.0 σ is drawn in green net and the anomalous density contoured at 2.5 s is drawn in blue net.

previously by Howerton and coworkers (Howerton et al., 2001). The final DDD-Rb^+ structure contains one partially occupied Rb^+ ion in the minor groove at the AT-step of the A-tract. Figure 13 shows the final anomalous map contoured around the DDD-Rb^+ structure contoured at 2.5 σ. The anomalous peaks of the phosphorous atoms are visible along the backbone. At the P3 site, the anomalous peak of the Rb^+ can be seen. In the final anomalous map, the phosphorous peaks ranged from 7.97 to 15.11 σ and the Rb^+ peak had an intensity of 8.95 σ. One phosphorous peak was less intense than the Rb^+ peak.

Anomalous map phasing.

Initial phases for the anomalous map were calculated using the positions of DNA phosphorous atoms only. Phosphorous positions were determined by refining the DNA model against the data. This method does not introduce any model bias in terms of Rb^+ positions. The phosphorous atoms alone should give essentially the same phases and the

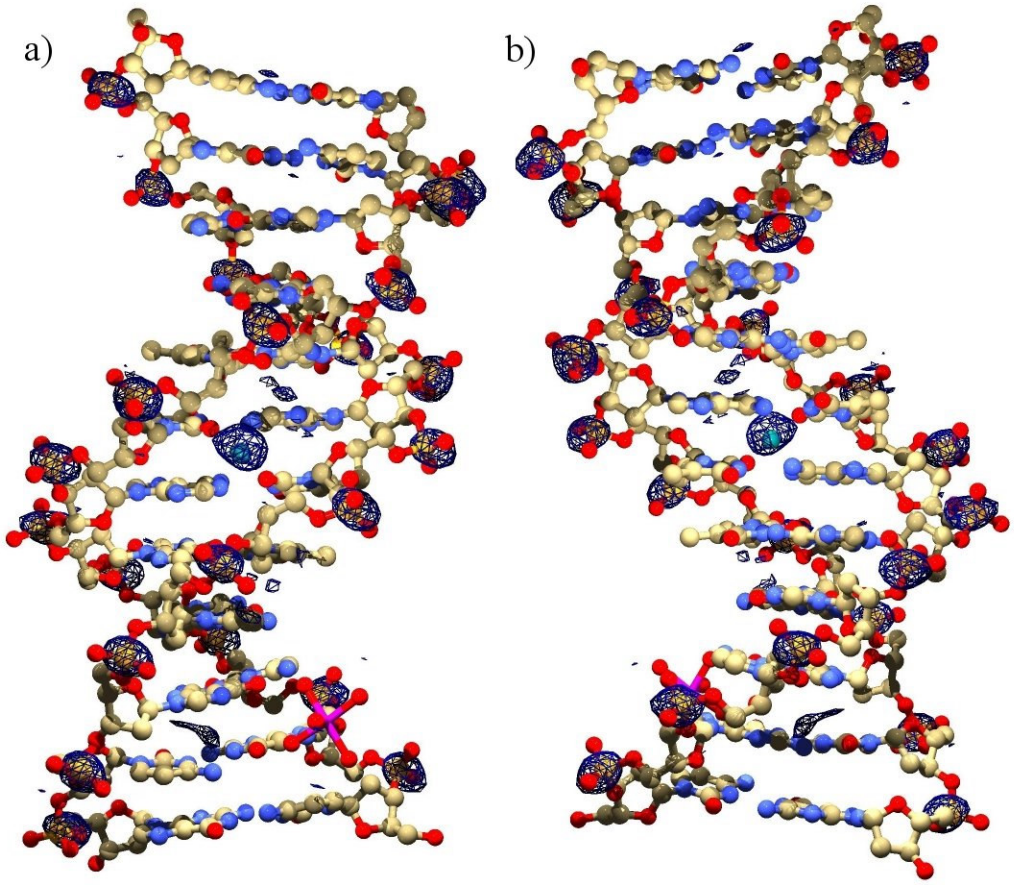

Figure 13. Anomalous density contoured at 2.5 σ around DDD-Rb^+.
Anomalous density (blue net) contoured at 2.5 σ around DDD-Rb^+. The Rb^+ at the P3 site is drawn in cyan and the Mg^{2+} in magenta. The figure shows the views minor (a) and major (b) grooves.

same maps as all DNA atoms, because f' and f'' of N, C, and O are significantly less than phosphorous (Table 3, Chapter II). N, C, and O atoms should not make a significant contribution to the phases. For reasons that are not clear, maps calculated with all N, C, O and P atoms were of poorer quality than maps calculated with P only. The measure of map quality is the σ level of the phosphorous anomalous peaks, which were less when all DNA atoms were used for phasing.

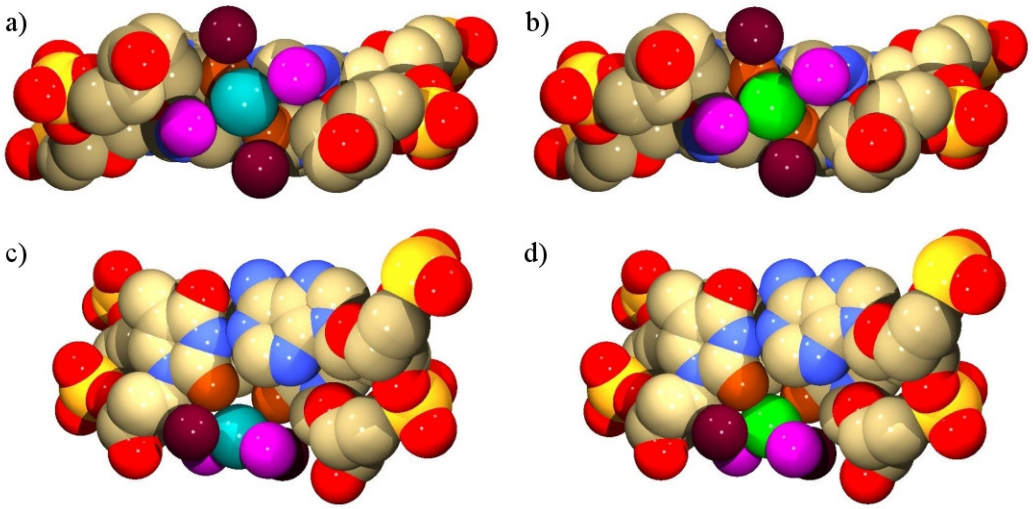

Figure 14. Comparison of Rb^+ and Tl^+ binding at the P3 site.
Details of Rb^+ (a and c) and Tl^+ (b and d) shown from the front (a and b) and top (c and d). Tl^+ (green spheres) makes a closer contact with O2 atoms (orange) of T7 and T19 than Rb^+ (cyan spheres). The O4' oxygen atoms of T8 and T20 are drawn in dark red spheres. The two water molecules coordinating the cations are drawn in magenta spheres.

Comparison of the P3 sites of DDD-Tl^+ and DDD-Rb^+.

The DDD-Rb^+ and DDD-Tl^+ structures were superimposed. An RMSD of 0.74 Å of atomic positions was calculated using CNS 1.1. When the four DNA oxygen atoms involved in binding to Tl^+ and Rb^+ ions are superimposed, it can be seen that the position of the Tl^+ ion is closer (by 0.53 Å) to the DNA then the position of the Rb^+ ion (Figure 14).

Valence calculations.

We compare the specific valences of Na^+, K^+, Rb^+ and Tl^+ at the P3 site. The positions of Rb^+ and Tl^+ in the P3 cite have been determined by refinement of models against data. In these cases the models contained explicit ions at the P3 site. However for Na^+ and K^+ the models contain water only at the P3 site. Therefore the Na^+ and K^+

Table 8. Calculated specific valences, occupancies, and chemical properties.

	Na^{+} [a]	K^{+} [a]	Rb^{+} [b]	Tl^{+} [b]
R_0 (Å)	1.6	2.3	2.2	2.1
N	4.3	9.1	7.0	6.0
R_j (Å)				
T19 O2	2.8	2.9	3.0	2.8
T7 O2	2.7	2.8	3.0	2.9
T20 O4'	3.1	3.4	3.3	3.1
T8 O4'	3.3	3.4	3.3	3.6
H_2O (36 or 3015[c])O	2.8	2.9	2.9	3.0
H_2O (53 or 3024[c])O	3.0	2.9	3.0	2.7
υ	0.49	0.55	0.60	0.81
Occupancy	0.30-0.50[d]	0.51±0.02[e]	0.53	0.10[c]
Chemical Properties				
Z	10	18	36	81
Ionic Radius (Å)[f]	0.95	1.33	1.48	1.49
ΔH_{hyd} (kJ/mol)[g]	-405	-321	-296	-326
Coordination Number[f]	6-7	4-7	>6	6-8
Hard/Soft Character[f]	Hard	Hard	Hard	Intermediate
Preferred Ligand[f]	O	O	O	O,N,S

[a]R_j measured from the H_2O at the AT-step to the coordinating oxygen atoms. [b]R_j measured from the cation at the AT-step to the coordinating oxygen atoms. [c](Hershkovitz et al., 2003) [d](Shui et al., 1998a) [e](Cesare Marincola et al., 2004) [f](Brown, 1988) [g](Wulfsberg, 1991)

positions are less certain than the Rb^{+} and Tl^{+} positions. The rubidium and thallium specific valences (υ_{Rb} and υ_{Tl}) were calculated with the equation:

$$\upsilon = \Sigma(R_j/R_0)^{-N}$$

as described by Nayal and DiCera (Nayal & DiCera, 1996). R_j is the measured distance between the cation and oxygen ligand j. R_0 is the length of a bond of unit valence, and N is a parameter obtained by fitting large number of complexes. The Rb^{+} and Tl^{+} specific valences (υ_{Rb} and υ_{Tl}) along with Na^{+} and K^{+} specific valences (υ_{Na} and υ_{K}) (reported previously in (Shui et al., 1998a)) are listed in Table 8, along with R_0 and N, established by Brown (Brown, 1981). For υ_{Na} and υ_{K} the position of the water molecule at the P3

site used to determine the R_j values. The three group I cations had specific valences ranging from 0.48 to 0.60. The highest ion-specific valence was that of thallium, 0.81.

Discussion

Rb^+ and Tl^+ have been used to probe DNA for cation localization (Howerton et al., 2003; Howerton et al., 2001; Tereshko et al., 1999; Tereshko et al., 2001). Rb^+ and Tl^+ have the advantage of strong x-ray scattering, and strong anomalous signals, providing a signal that allows discrimination from water. The assumption used in the Rb^+ and Tl^+ substitution experiments is that Na^+, K^+, Rb^+, and Tl^+ are similar in their chemical properties and bind to similar sites on DNA.

Here we have analyzed the X-ray structure of the DDD crystallized in the presence of Rb^+ and compared it to the structure of the same dodecamer in the presence of Tl^+. Egli and coworkers have previously determined several structures of the DDD obtained from crystals grown in the presence of Rb^+ (Tereshko et al., 1999). The highest resolution of those structures is called DDD^E-Rb^+ here. Unfortunately, DDD^E-Rb^+ cannot be directly compared to DDD-Tl^+ because (i) the DNA of DDD^E-Rb^+ is covalently modified by 2' fluoro groups, (ii) the crystallization solution of DDD^E-Rb^+ contained significantly higher [Mg^{2+}] (25 mM) than that of DDD-Tl^+ (5.2 mM), (iii) the final refined model of DDD^E-Rb^+ contains three Mg^{2+} ions while DDD-Tl^+ contains one Mg^{2+}, (iv) the crystallization solution of DDD^E-Rb^+ contained lower spermine (3.2 mM) than that of DDD-Tl^+, (8.9 mM), (v) the final refined model of DDD^E-Rb^+ lacks spermine molecules while DDD-Tl^+ contains one partially ordered spermine molecule, and (vi) Rb^+

ions in the DDD^{E}-Rb^{+} were identified by difference Fourier methods while Tl^{+} ions in the DDD-Tl^{+} were identified by anomalous methods.

The experiment conditions and methods of DDD-Tl^{+} were replicated as closely as possible here for DDD-Rb^{+}. The key differences are that DDD-Tl^{+} crystals were grown from solutions of approximately half the Mg^{2+} concentration and over twice the heavy monovalent cation concentration as DDD-Rb^{+}. These differences were imposed by the requirement for highly quality crystals. From these differences in experimental conditions it was anticipated the observed occupancies of Rb^{+} and the number of Rb^{+} sites might be less than those observed with Tl^{+}. In addition, it merits mention that the anomalous signal of Tl^{+} (f'=-4.881, f''=8.085 at 1.54 Å) is significantly greater than that of Rb^{+} (f'=-0.574, f''=1.608 at 1.54 Å) at the wavelength used for data collection. For low occupancy sites Tl^{+} is a much more sensitive probe than Rb^{+}.

Similarities in cation distributions of DDD-Tl^{+} and DDD-Rb^{+}.

There is one over-riding element of similarity between the cation distributions of DDD-Tl^{+} and DDD-Rb^{+}. Both structures show cation localization in the minor groove of the A-tract. The common site of localization is adjacent to the ApT-step, and was previously termed the P3 site (Shui et al., 1998a; Shui et al., 1998b). The P3 site contains an Rb^{+} ion in DDD^{E}-Rb^{+}, a K^{+} (Sines et al., 2000) or Cs^{+} (Woods et al., 2000) when those cations are contained in the crystallization solution. Therefore, it is beyond dispute that in crystal structures of the DDD, the P3 site localizes a variety of monovalent cations, and is the preferential monovalent cation binding site under a variety of conditions.

Differences in the position and occupancy of the cation at the P3 site.

In the P3 site, the positions of Tl^+ and Rb^+ are slightly different. Tl^+ resides 0.53 Å closer than Rb^+ to the floor of the groove. This difference is observed even though the positions of the DNA oxygen ligands *relative to each other, and relative to the floor of the groove* are conserved. The Tl^+ ion appears to be in very tight contact with coordinating keto oxygens (O2 atoms of thymines). These short distances may be suggestive of the non-spherical nature of the Tl^+ ion, discussed below.

The estimated Rb^+ occupancy in the P3 site (0.52) is significantly greater than that of Tl^+ (0.10). Although the occupancies are rough estimates obtained by manual iterative minimization of positive and difference electron density, we believe the difference between Rb^+ and Tl^+ occupancies is real and significant. In fact the occupancy of Rb^+ in DDD-Rb^+ is consistent with that obtained previously in DDD^E-Rb^+.

Additional cation sites.

In addition to the P3 site, twelve Tl^+ sites were identified in DDD-Tl^+. None of non-P3 sites observed with Tl^+ are observed in the DDD-Rb^+ anomalous map. Figure 15 shows the DDD-Rb^+ anomalous map in the region of highest Tl^+ occupancy in (in DDD-Tl^+). In this region there is a poor correspondence between the Rb^+ and Tl^+ anomalous maps. Considering the lower sensitivity of Rb^+ compared to Tl^+, one cannot conclude that none of the Tl^+ sites are correspondingly occupied by Rb^+. We estimate that with our experimental conditions a Rb^+ occupancy of at least around 20% would be required to give a clear peak (1-2 σ) in the anomalous map. By contrast, an occupancy of 10% Tl^+ gives a clear peak in the anomalous maps. Probably the only safe conclusion is that the preferences of Rb^+ and Tl^+ are different. The highest occupied Rb^+ site is in the minor

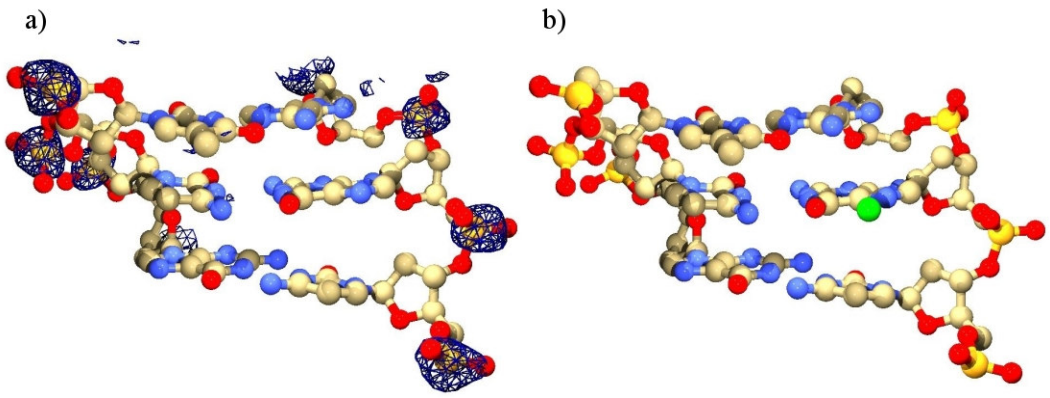

Figure 15. Comparison of the highest occupancy Tl^+ site in DDD-Tl^+ with DDD-Rb^+. Comparison of a sample cation binding site between DDD-Rb^+ and DDD-Tl^+. DDD-Rb^+ (a) is shown with the anomalous density contoured at 2.5 σ drawn in blue net. DDD-Tl^+ (b) is shown without the anomalous map and with the assigned Tl^+ ion at the same site in the major groove

groove, at the P3 site, whereas the highest occupancy Tl^+ sites are in the major groove. The P3 site ranks eighth in Tl^+ occupancy.

Ion-specificity of coordination geometry at the P3 site.

It is of interest to know if cation occupancies in the DDD ensemble are modulated by cation-specific coordination geometry. The experimentally-determined occupancies of Na^+, K^+, Rb^+ and Tl^+ at the P3 site may be influence by ligand geometry. For example, the greater occupancy of Rb^+ relative to Tl^+ might originate from better Rb^+ coordination geometry. This issue is worthy of attention even though the occupancies are rough estimates, and are influenced by differences in crystallization conditions, etc. The estimated occupancies are

Rb^+ (0.53, here) > Na^+ (0.51, (Cesare Marincola et al., 2004)) >

K^+ (0.35, (Sines et al., 2000)) > Tl^+ (0.10, Howerton et al., 2001).

Neither we nor others have obtained reliable estimates of Na^+ occupancy at specific sites in DNA crystals. The contrast with water is insufficient. In the absence of

diffraction information one can look to solution ^{23}Na and ^{87}Rb magnetic relaxation dispersion (MRD) experiments (Cesare Marincola et al., 2004; Denisov & Halle, 2000). Those experiments suggest an Na^+ occupancy of 0.51 ± 0.02. However the solution NMR experiment is conducted with a different set of conditions (solution versus crystal) and furthermore does not provide information on specific location of the Na^+ (in P3 versus other minor groove sites). Therefore the Na^+ occupancy should be considered the least reliable occupancy of the series.

An ion-specific valence calculation (Brown, 1981; Brown, 1992; Brown, 1996) gives an estimate of deviation from optimum metal-ligand geometry. A valence value of 1.0 indicates optimal geometry. A valence of less than 1.0 indicate that metal to ligand distance are longer or fewer in number than optimum. We compare the ion-specific valences of Na^+, K^+, Rb^+ and Tl^+ at the P3 site. The goal is to determine if the valences correlate with occupancies. The ion specific valences are

$$Tl^+ (0.81) > Rb^+ (0.60) > K^+ (0.55) > Na^+ (0.49).$$

There does not appear to be a clear relationship between occupancy and valence. For example Tl^+ has the highest valence, and the lowest occupancy.

What remains to be addressed is the reason why valence is not a good predictor of occupancy. The failure of the valence model may point to fundamental differences in coordination chemistry of Tl^+ and group I metals. Table 8 contains some relevant chemical properties of Na^+, K^+, Rb^+, and Tl^+. Tl^+ has an ionic radius and enthalpy of hydration very similar to those of K^+. However Tl^+ is non-spherical (Kristiansson, 2002) and is intermediate between hard and soft metal (Wiesbrock & Schmidbaur, 2003b) and

has affinity for soft and hard electron donors, although it has higher affinity for soft metals than hard metals.

In contrast, the group I metals are classified as hard metals. Tl^+ also exhibits a variety of coordination geometries with coordination numbers of 6 to 8 (Brown, 1988). In several crystals of Tl^+ and various organic salts, the Tl^+ is 'half naked', showing low coordination number, with one hemisphere unoccupied even in excess of available ligands (Janiak, 1997; Kristiansson, 2002; Wiesbrock & Schmidbaur, 2003a). Group I metals on the other hand are observed in geometries that maximize their coordination number and occupy their entire coordination sphere. Tl^+ forms Tl-O and Tl-N bonds like the group I metals, but it can also participate in Tl-π bonds with aromatic ring systems and long distance Tl-Tl "thallophilic" bonds. Thallium's non-spherical outer shell and its unusual coordination characteristics are often attributed to its inert partially hybridized $6s^2$ pair of electrons. The $6s^2$ orbital is contracted due to stronger than expected attraction between the electrons in this orbital and the nucleus.

In crystallographic studies of complexes between Tl^+ and benzoate, anthranilate and salicylate ligands, the cation geometry is described as a distorted pyramid with all the ligands at the base of the pyramid and the $6s^2$ lone pair at the apex pointing into a hydrophobic environment. The lone pair commonly interacts with the π electrons of aromatic ligands. In structures where long Tl-Tl bonds are observed, these thallophilic contacts seem to be weak and have very little effect in crystal structure determination. The length of this type of bond is often near the sum of the van der Waals radii of the two Tl^+ cations or slightly longer. Except for the Tl-Tl bonds, all the other modes of Tl^+ bonding to ligands are seen the 13 Tl^+ ions in DDD-Tl^+, hence indicating that Tl^+

interactions with B-DNA have the same characteristics as Tl^+ interactions with inorganic ligands. Two prominent examples of this are thalliums 2111 and 2112, which are coordinated in the intermolecular space between adjacent DNA molecules. Both of these Tl^+ ions are coordinated to the π electrons of cytosines (Howerton et al., 2001).

The distorted square pyramidal geometry of the Tl^+ cation may also explain the higher affinity of the P3 site for Rb^+ than Tl^+. At the P3 site the O2 atoms of T7 and T19 coordinate to one hemisphere of the cation. The O4' of the T8 and T20, and the oxygens of two water molecules from the second layer of the spine of hydration coordinate to the cation from the opposite hemisphere. This geometry is not consistent with the coordination geometry of Tl^+ seen in its crystal structures with benzoate and anthranilate ligands. In contrast, group I metals prefer to have their entire coordination sphere occupied by ligands. Therefore, the P3 site may be more suitable for Na^+, K^+, and Rb^+ than for Tl^+.

Although there are clear differences between Tl^+ and the group I metals, it must be noted that in many systems Tl^+ is a reasonable mimic for Na^+ and K^+. Tl^+ and K^+ can bind to and stabilize DNA G-quartets (Basu et al., 2000). Tl^+ can effectively substitute for K^+ in the sodium-potassium pump (Pedersen et al., 1998), in activation of pyruvate kinase (Loria & Nowak, 1998), and in catalytic activity of fructose-1,6-bisphosphate (Villeret et al., 1995). Tl^+ has been used as an effective probe for binding of K^+ in X-ray crystal structures of B-form DNA (Howerton et al., 2001), DNA-drug complexes (Howerton et al., 2003), Tetrahymena ribozyme P4-P6 domain (Basu et al., 1998) and insulin (Badger et al., 1994a; Badger et al., 1994b; Gursky et al., 1992a; Gursky et al.,

1992b). The degree to which Tl^+ can mimic the group I monovalent cations will depend on the geometry of the coordinating ligands available for binding.

CHAPTER V

STRUCTURE OF B-DNA WITH CATIONS TETHERED IN THE MAJOR GROOVE

Abstract

The x-ray crystal structure of the dodecamer $[d(CGCGAAXXCGCG)]_2$ (X = Z3dU, 5-(3-aminopropyl)-2'-deoxyuridine) was determined at pH of 6.7 and in presence of thallium (I) acetate. At this pH, the amino group of the modification carries a 1^+ charge. This structure contains a total of four modifications in the major groove, two partial spermine molecules and eight partially occupied thallium (I) cations. One of the aminopropyl modifications (Z3dU 20) extends from the major groove to interact with the phosphates of the complementary strand. This results in changes in inter-basepair rise, roll, and helical twist, and ultimately, deformation of the backbone.

All four modifications extend in the 3'-direction. Aside from the tethered amino group of Z3dU 20, the other three modifications extend into the intermolecular cavities of the crystal. Additionally, eight Tl^+ sites are identified in the structure using the anomalous scattering of thallium. All but one of the Tl^+ sites are different from the 13 Tl^+ sites identified previously (Howerton et al., 2001) in the x-ray crystal structure of the unmodified dodecamer. This site is the P3 site (ApT-step), where previously many cations have been shown to bind.

Analysis of the monovalent cation distribution and comparison with previous findings provided evidence for the redistribution of cations as a result of Z3dU 20 partial interaction with the backbone. Specifically, Tl^+ ions appear to be ejected from the

vicinity of this tethered cation in the major groove. This chapter is based on a manuscript prepared for publication in collaboration with Tatsuya Maehigashi, Derrick Watkins, Dr. Seiji Komeda, George T. Lountos, who kindly collected the data, and Dr. Barry Gold, Dr. Luis Marky, Dr. Micheal Stone and Dr. Loren D. Williams.

Introduction

DNA conformation is influenced by a variety of factors. In a ground state structure (i) interactions of water molecules with the hydrophobic planer surfaces of DNA bases are minimized, (ii) stacking and hydrogen-bonding interactions are maximized, and (iii) intra-molecular phosphate-phosphate repulsions are minimized. It is empirically known that DNA sequence is an important modulator of conformation. A-tracts are associated with narrow minor grooves (Alexeev et al., 1987; Burkhoff & Tullius, 1987; Wing et al., 1980), axial bends (Diekmann & Wang, 1985; Hagerman, 1984; Marini et al., 1982), and high propeller twists (Drew et al., 1981). G-tracts exhibit strong divalent cation dependence of bending (Brukner et al., 1994) and a propensity to convert to A-form (Hud & Plavec, 2003). The sequence-dependence of DNA conformation can be ascribed to base-specific hydrophobicity, shape, charge distribution, and functional group disposition. The relative importance and mechanisms by which various sequence-specific factors influence DNA conformation is not currently resolved.

Here we examine the X-ray structure of the DDD (duplex Dickerson dodecamer) covalently tethered to four cations. This modified version of the DDD, called here the $^{XX}DDD^{4+}$ is composed of $[d(CGCGAAXXCGCG)]_2$ where X is effectively a thymine residue linked at the 5-position to an n-propyl-amine (Z3dU, Figure 3, Chapter II). The

four n-propyl-amino groups of $^{XX}DDD^{4+}$ are protonated under the conditions of our experiments. The 4+ of $^{XX}DDD^{4+}$ represents the charge of the four modifications. The net charge of the $^{XX}DDD^{4+}$, including contributions from phosphate groups, is 18^-.

The goal of the work here is to determine the effects of tethering cationic charge to DNA on conformation, and on distribution of non-tethered counterions such as K^+, Mg^{2+} and spermine. To determine non-tethered cation positions the structure of $^{XX}DDD^{4+}$ was determined in the presence of Tl^+ ions. Thallium I, with a strong anomalous signal, has been broadly used as a mimic of K^+ in x-ray diffraction experiments aimed at determining positions of cations adjacent nucleic acids (Basu et al., 1998; Howerton et al., 2003; Howerton et al., 2001) and proteins (Badger et al., 1994a; Badger et al., 1994b; Brown, 1988; Gill & Eisenberg, 2001; Gursky et al., 1992a; Gursky et al., 1992b; Loria & Nowak, 1998; Pedersen et al., 1998; Stroud et al., 1974; Villeret et al., 1995). In previous work from our laboratory, 13 Tl^+ sites were identified in the solvent region of the unmodified form of the DDD-Tl^+ (native DDD) (Howerton et al., 2001). Here we present evidence that tethering of cations into the major groove influences DNA deformation and the distribution of the counterion Tl^+.

The motivation for this work stems ultimately from work of Mirzabeckov, Rich and Manning (Manning et al., 1989; Mirzabekov & Rich, 1979), who considered the proposition that charge neutralization of phosphate groups might drive DNA bending. They were concerned with asymmetric charge neutralization by cationic protein sidechains, leading to 'phosphate collapse'. That model was extended by Rouzina and Bloomfield (Rouzina & Bloomfield, 1998), who proposed that mobile cations such as Mg^{2+} can contribute electrostatically to DNA bending. In the Rouzina and Bloomfield

model, a divalent cation can repel and displace other counterions, leading to strong attraction with unscreened phosphate groups. The result is DNA bending by electrostatic collapse around a divalent cation. Our experimental system has limitations for analysis of DNA bending. However, it does provided a detailed view of microheterogenity in DNA conformation.

Materials and Methods

Crystallization.

Reverse-phase HPLC purified d(CGCGAAXXCGCG) was synthesized and purified as described (Heystek et al., 1998; Li et al., 2002). The oligomer was annealed by slow cooling from 80ºC. Crystals were grown over 2-3 days using the sitting drop vapor diffusion method. The crystallization solution contained 0.6 mM DNA (duplex), 20 mM spermine•cacodylate (pH 6.5), 4.0% v/v 2-methyl-2,4-pentanediol (MPD), 4.0 mM spermine acetate (pH 6.5), and 20% (v/v) ANAPOE™-$C_{12}E_{10}$ (Hampton Research). DNA concentration was determined by UV spectroscopy. The crystallization solutions were equilibrated against a reservoir of 35% v/v MPD at 22ºC. The spermine•cacodylate solution was prepared by titrating a solution of cacodylic acid (Sigma) with a solution of spermine base (Sigma). The concentration of spermine in the initial crystallization solution, including contributions from the spermine•cacodylate solution, was 8.0 mM. Efforts to grow crystals of dCGCGAAXXCGCG in the presence of Tl^+ were not successful. A wide range of pH, Mg^{2+}, spermine, and Tl^+ concentrations were assayed.

It was concluded that dCGCGAAXXCGCG shows a propensity to crystallize only under low cationic strength. The conditions reported here are at the maximum limit

of cation concentration that produces diffraction quality crystals. Since it was not possible to grow high quality crystals of dCGCGAAXXCGCG from solutions containing sufficient Tl^+, crystals were soaked with Tl^+. The initial soaking solution contained 40 mM thallium (I) acetate, 20mM spermine•cacodylate (pH 6.5), 4.0% v/v MPD, 4 mM spermine acetate (pH 6.5), and 20% v/v Anapoe™-$C_{10}E_{12}$. Equivalent volumes of crystallization and the initial soaking solutions were equilibrated against a reservoir of 35% v/v MPD and combined to give the final soaking solution. A crystal of dimensions 0.1 x 0.1 x 0.1 mm^3 was transferred to the final soaking solution. After 6 hours the crystal was looped and frozen in liquid nitrogen.

Data collection and processing.

360º of data, with oscillation angle of 1.0º was collected at beamline 22-ID in the SER-CAT facilities at APS using a MAR-165 CCD detector at a wavelength of 1.00930 Å. The crystal was maintained at -160ºC. Data were scaled and integrated using HKL2000 version 1.97.7 (Otwinowski, 1993). Initial inspection of the images suggested twinning. The quality of the data was significantly improved by excluding a subset of frames from scaling based on HKL2000 statistics. Frames 16-20, and 126-130 of the 360 frames were excluded from the scaling process. A total of 146,309 reflections were indexed, integrated, and reduced to 8901 unique reflections with merged Bijoet pairs and 15,840 unique reflections with unmerged Bijoet pairs.

Unit cell dimensions are a = 24.99 Å, b = 41.06 Å, and c = 66.34 Å, in space group $P2_12_12_1$. The data used in the refinement included 6586 unique reflections from 75 to 1.51 Å. Table 9 contains the data collection and refinement statistics.

Table 9. Crystallographic and refinement Statistics for $^{XX}DDD^{4+}$.	
unit cell	
a (Å)	24.99
b (Å)	41.06
c (Å)	66.34
space group	$P2_12_12_1$
temp. of data collection (°C)	-160.0
No. of reflections	146309
No. of unique reflections	8901
completeness (%)/highest shell (%)	77.3/49.1
max resolution of obsd reflections (Å)	1.49
max resolution of highest shell used (Å)	1.60
resolution range (Å)	35-1.60
No. of reflections used in refinement	6586
No. of reflections used in the test set	769
rmsd of bonds from ideal (Å)	0.0097
rmsd of angles from ideal (Å)	1.3401
DNA (asymmetric unit)	$[d(CGCGAAXXCGCG)]_2$
No. of DNA atoms	498
No. of water molecules	114
No. of Mg^{2+} ions	0
No. of spermine atoms	15 (2 partial molecules)
R-free (%)	25.24
R-factor (%) excluding test set data	19.73
final *R*-factor (%) using all data	23.33
unit cell volume occupied (%)	61.25

Refinement.

Molecular replacement was used for phase determination. The coordinates for the starting DNA model were obtained from NDB (entry BDL084) (Shui et al., 1998a). The program CNS version 1.1 (Brunger et al., 1998)) was used, along with the parameters of Berman and coworkers (Clowney et al., 1996; Gelbin et al., 1996; Parkinson et al., 1996).

The parameter files were modified to remove dihedral restraints, except for the addition of dihedral restraints to keep the planarity of purine and pyrimidine ring systems. Restraints for Z3dU, estimated by combining elements of thymine and lysine, were incorporated into the parameter files.

Initial electron density maps and refinement statistics from starting model BDL084 indicated an incorrect solution. The structure was solved by a rotational/translational search and rigid body refinement in CNS. At this stage, the position of the Z3dU modifications could be identified in the difference electron density maps. The four thymine residues in structure BDL084 were converted to Z3dU. The resulting model was annealed and refined. Water molecules were added iteratively followed by annealing, refinement, and phase calculation using peaks in the sum ($2|F_o|-|F_c|$) and difference ($|F_o|-|F_c|$) maps.

Anomalous ($|F^+|-|F^-|$) maps showed the positions of Tl^+ ions. The initial phases for the anomalous map were obtained from the positions of DNA phosphorous atoms. Six possible Tl^+ sites were identified in this map. Water molecules previously located at these sites were converted to partially occupied Tl^+ atoms. Successive refinements and map calculations were performed. Estimates of occupancies were obtained by monitoring negative and positive difference electron density. Additional anomalous maps were made using the position of the phosphorous atoms and the six partially occupied Tl^+ atoms to calculate phases. Three additional Tl^+ sites were thus identified. These sites where added to the model in the same manner as the first six; however, only two of these sites resulted in acceptable geometry and statistics. Helical parameters were calculated with the program Curves 5.3 (Stofer & Lavery, 1994).

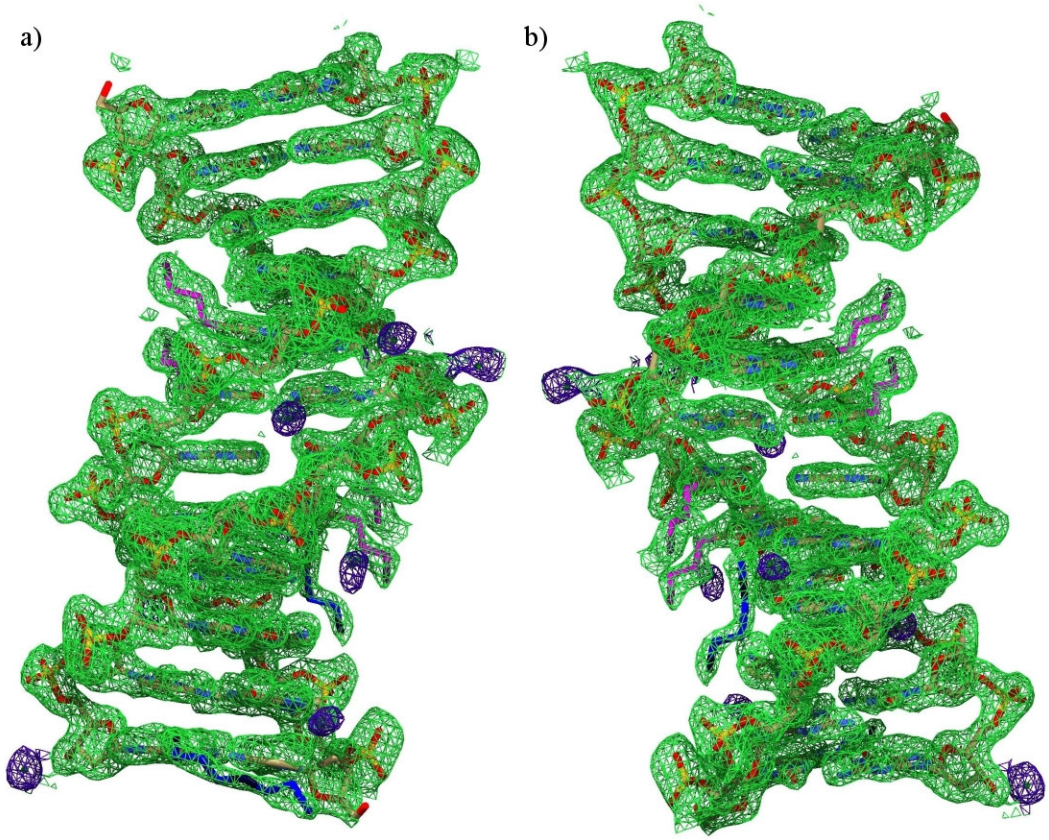

Figure 16. Sum electron density contoured at 1.0 σ around $^{XX}DDD^{4+}$.
Sum density is drawn in green net around DNA atoms and in blue net around Tl^{+} ions. The modifications are in magenta with the terminal N5D in dark magenta. Spermine molecules are in blue with the nitrogen atoms in dark blue. The figure shows the view of minor (a) and major (b) grooves.

Results

DNA conformation.

The final model of DDD^{4+}, including the amino-propyl groups, is well-determined by the data (Figure 16). The occupancies of the two terminal O5' atoms of the DNA were set to zero because they appear to be disordered; there is no electron density observed around them.

Each of the four amino-propyl modifications is readily identifiable in sum and difference maps (Figure 17). Three amino-propyl groups (on residues 7, 8, and 19)

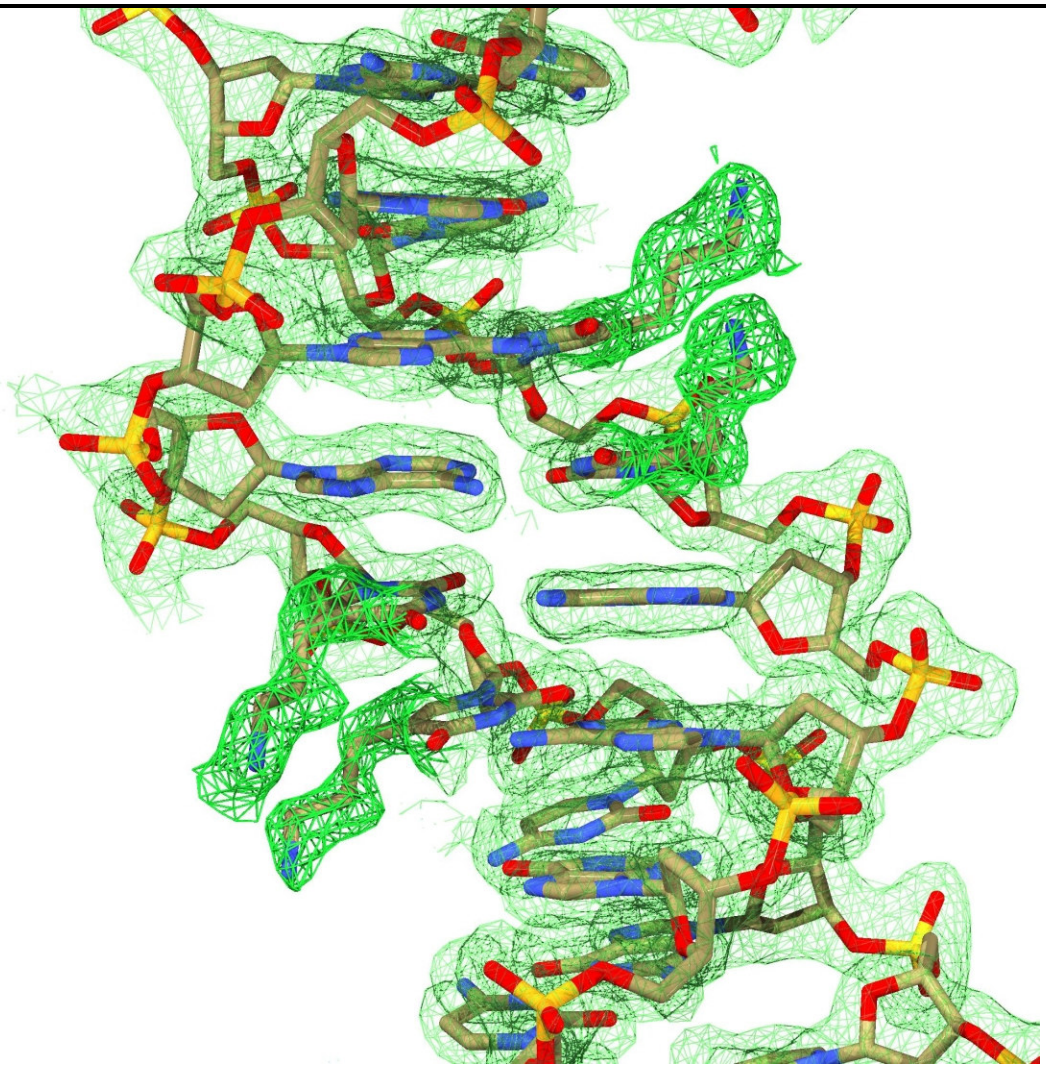

Figure 17. Detailed view of sum electron density around the 4 Z3dU in $^{XX}DDD^{4+}$.
Sum electron density contoured at 1.0 σ drawn around XXDDD4^{+}. The view is from the major groove. The density is drawn in green net around the ω-aminoproply modifications of the four Z3dU residues and in transparent (very light) green, elsewhere.

extend radially out from the DNA, and into intermolecular cavities in the crystal. These cationic amino groups do not appear to engage in significant interactions with the DNA backbone or bases.

The relative strength of various electrostatic forces between N^{+} (N^{+} refers to the cationic amino nitrogen of the aminopropyl of Z3dU) and phosphate oxygens (OP) of the

DNA are indicated symbolically in Figure 18. Each circle in this figure represents a single N^+-OP distance. All distances less than 10 Å are indicated. The sizes of the circles are proportional to $1/r^2$, where r is the N^+ to OP distance. The blue and red circles indicate intra-duplex distances. These interactions would be observed in an isolated duplex in solution, assuming the conformation in crystal and solution is conserved.

The interactions of the amino-propyl modification of residue 20 are indicated by red circles in Figure 18. The amino-propyl modification of residue 20 extends partially in the 3' direction. The N^+ of residue 20 ($20N^+$) appears to be involved in significant electrostatic interactions with the DNA. These circles are focused primarily at one end of the duplex. The phosphates on this end of the duplex are engaged in significant electrostatic interactions with the $20N^+$. The largest circle is adjacent to the 2-23 base pair, and indicates a distance of 6.4 Å. This interaction involves the O1P atom of guanosine 2.

The blue circles indicate intra-duplex distances of N^+ of Z3dU residues 7, 8 and 19. These distances are relatively long, and are evenly distributed over the duplex. A third set of circles (green) in Figure 18 represents interduplex interactions. These are 'lattice interactions' between adjacent duplexes in the crystal. These interactions are focused at the opposing end of the duplex. The largest circle there represents an interaction of 5.0 Å with the O2P atom of cystosine 15. Since these lattice interactions are focused primarily at the end of the duplex with conserved conservation, it can be argued that they are not inducing the observed deformation. The intra-duplex electrostatic interaction of $20N^+$ with the DNA appears to distort of the backbone and major groove geometry.

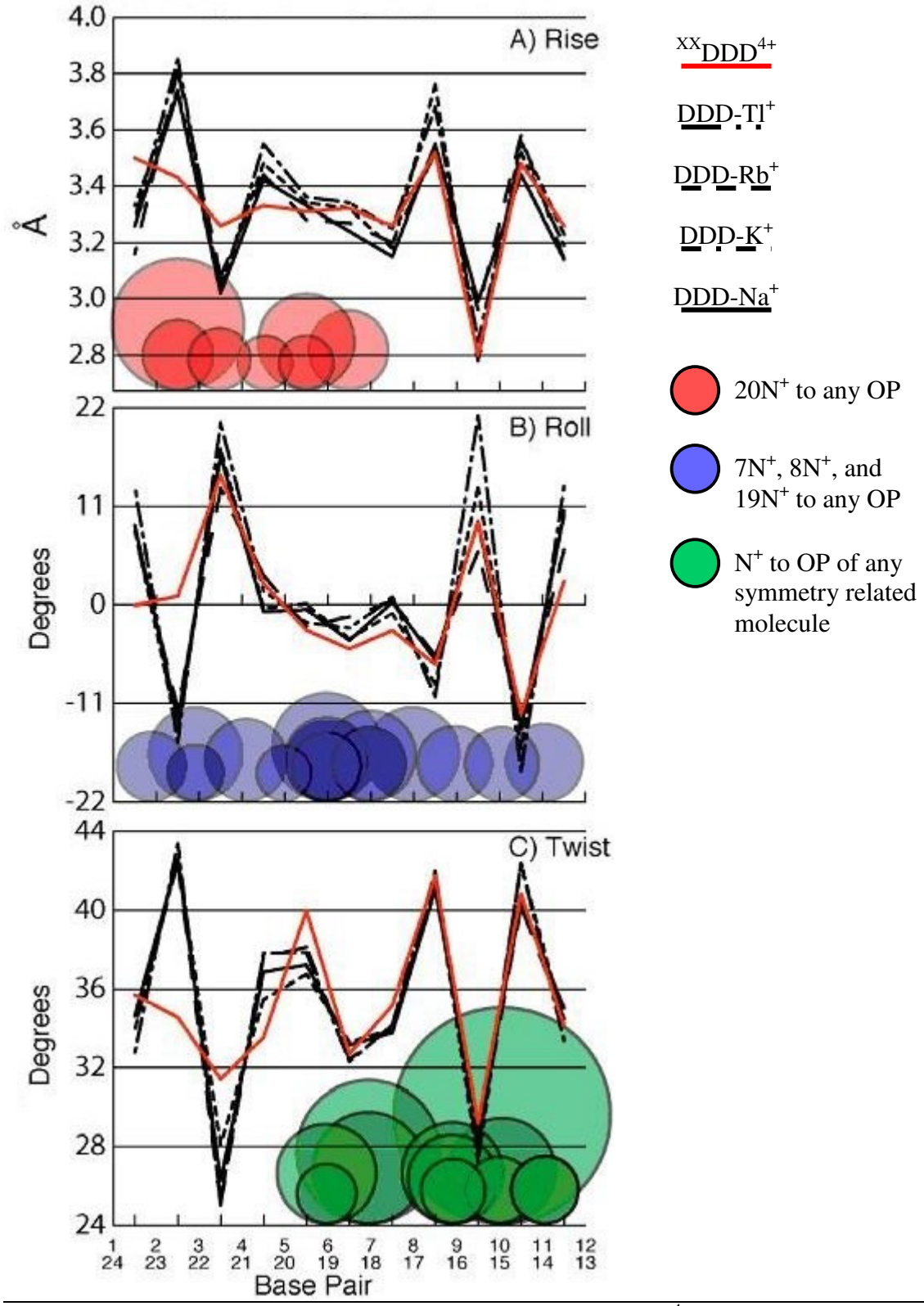

Figure 18. Comparison of rise, roll, helical twist between DDD^{4+} and native DDD.
All of the structures were obtained from crystals grown in presence of low magnesium concentration. The radii of the circles are scaled by $1/R^2$, where R is the N^+ to O1P/O2P (OP) distance.

These distortions are inferred by comparing DDD^{4+} with native DDD structures. Here we compare DDD^{4+} with a series of DDD structures determined from crystals obtained in the presence of various cations, in our laboratory. The members of this ensemble were determined independently and give an indication of the accuracy of the helical parameters. The helical parameters are shown in Figure 18. The rise, roll, and helical twist parameters of DDD^{4+} are different from the native DDD ensemble.

By this comparison (Figure 18), it can be seen that one half of the DDD^{4+}, from base pairs 6-19 to 12-13, is representative of the ensemble, and does not deviate significantly from it. However, the other half of DDD^{4+}, from base pairs 1-24 to 5-20 does deviate from the DDD ensemble. The deviations are localized primarily to base pairs 2-23 and 3-22. The helical rise of DDD^{4+} at the step from base pair 2-23 to base pair 3-22 is 3.4 Å, which is significantly less than the 3.7 Å rise observed for the ensemble. At the same step the roll of DDD^{4+} (0°) is over 10° greater than the ensemble (-11°). Similarly at this step the helical twist of the DDD^{4+} (34°) is much less than the ensemble (42°). The important point of this analysis is that the deviations of DDD^{4+} from the DDD ensemble are localized to the region where the interaction of tethered cation with the DNA phosphate groups is most significant.

A superimposition of DDD^{4+} with a native DDD structure is shown in Figure 19. The differences between the two DNA duplexes are limited primarily to the region of interaction of the amino group of ZdU 20 with DNA phosphate groups. The greatest atomic displacements of DDD^{4+} compared to DDD is in residues 1 and 2 in one strand and residues 21-24 in the complementary strand. The deviation between the two structures is greatest in backbone atoms. An asymmetric displacement of the two strands

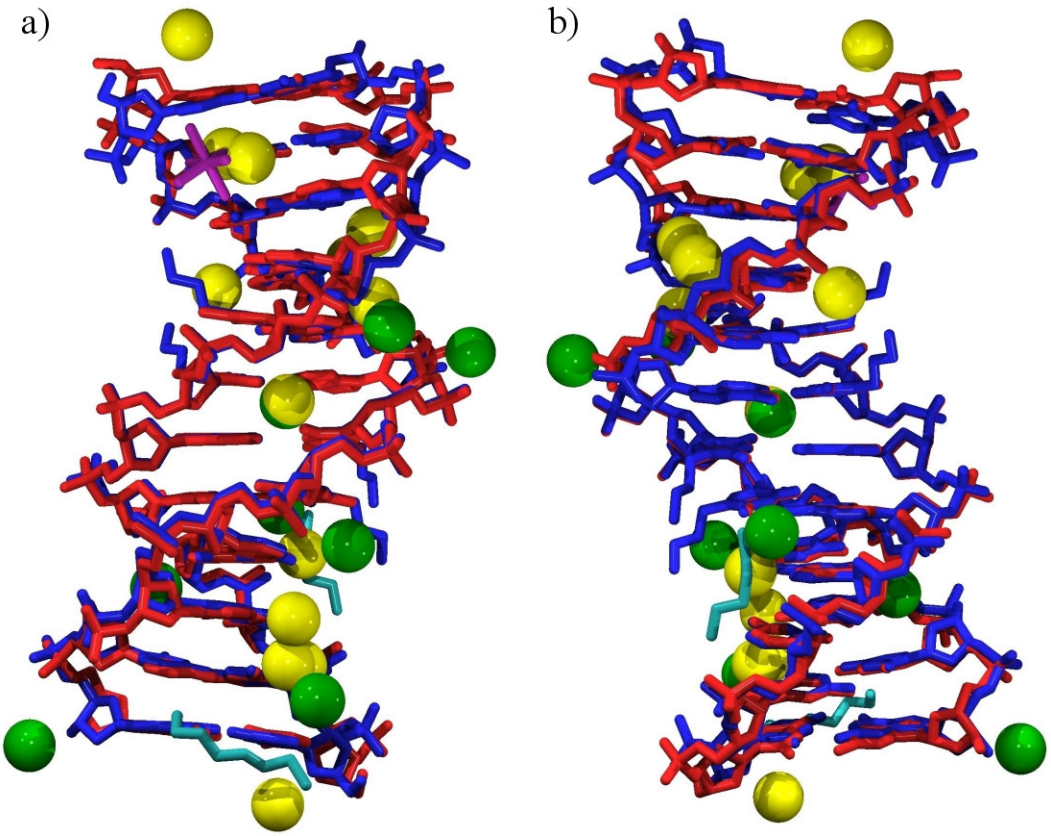

Figure 19. Superimposed structures of $^{XX}DDD^{4+}$ and DDD.
The superimposed structures of the $^{XX}DDD^{4+}$ (blue, with green Tl^{+} ions) and DDD-Tl^{+} (red, with yellow Tl^{+} ions). The corresponding spermines and $Mg^{2+}(H_2O)_6$ for each structure are in cyan and magenta, respectively. The figure shows the view of minor (a) and major (b) grooves.

has resulted in an asymmetric collapse of the major groove in this terminus. DDD^{4+} appears to be over-twisted. In addition, difference density near the O1P and O2P of G 2 suggests that this phosphate group is disordered. The phosphate appears to occupy a variety of positions along a trajectory in and out of the major groove. Figure 20 shows the electron density maps and details of the interaction between the tethered cation of Z3dU 20 and the backbone phosphates. Table 10 lists the relevant distance and angle information. A water molecule (H_2O 96) bridges $20N^{+}$ and O1P of G 2 and phosphate group of a symmetry-related duplex.

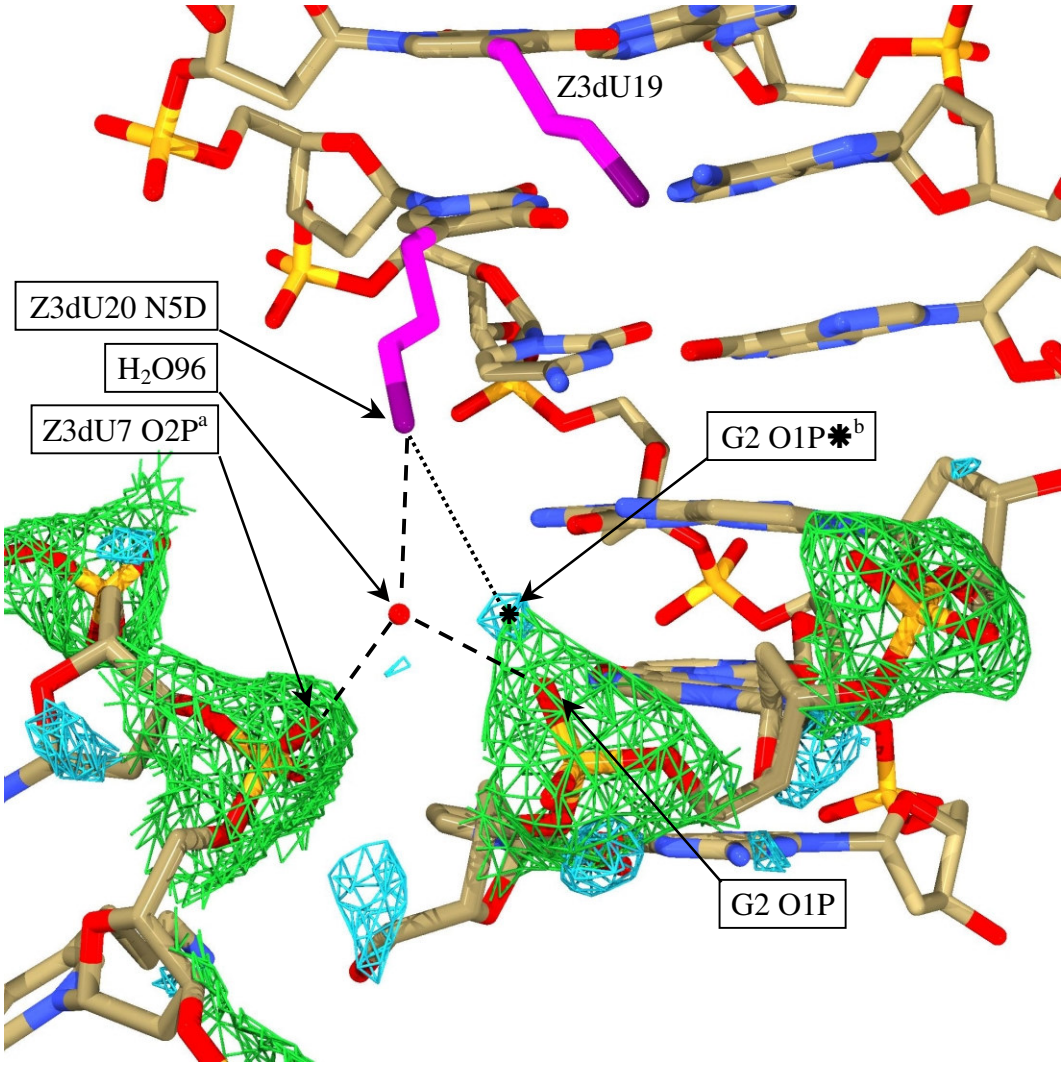

Figure 20. Detailed view of the contact between N5D of Z3dU 20 and the backbone.

Table 10. List of distances and angles for the relevant contacts with N5D of Z3dU 20.

Distances (Å)		Angles (degrees)	
N5D-H_2O	3.59	N5D-H_2O-O1P	111.95
N5D-O1P	6.40	N5D-H_2O-O2P	135.43
N5D-O1P✱	4.91	O1P-H_2O-O2P	84.75
N5D-O2P	6.14		
O1P-H_2O	4.13		
O2P-H_2O	3.05		

The sum electron density contoured at 1.0 σ (green net) and difference density contoured at 2.5 σ (cyan net) are drawn around the phosphates of the backbone of the same molecule and a symmetry related molecule (lower left corner). The two aminopropyl groups are drawn in magenta with N5D atoms in dark magenta. [a]O2P of Z3dU7 from a symmetry related duplex. [b]O1P✱ is the alternative position for O1P, where it is at its closest approach to N5d of Z3dU 20. This position is at the center of the difference peak near O1P.

At its closest approach, the distance between the O1P of G 2 and the N^+ is estimated to be 4.9 Å. This distance was estimated from the position of the N^+ atom and the center of the difference peak near the O1P of G 2. A single water molecule (H_2O 96) bridges between the terminal amino group and the O1P of G 2 and O2P of Z3dU 7 (from an adjacent duplex). This water molecule is 1.1 Å below the plane of N^+, O1P, and O2P. Therefore, it only partially screens N^+ from O1P of residue 2. Directly below this water molecule (2.4 Å) is another water molecule (H_2O 60). Although not modeled in this way, it is possible that H_2O 60 and H_2O 96 are partially occupied, and are effectively the same water molecule at two different positions. It seems likely that when the phosphate of G2 swings into the major groove, the bridging water molecule moves deeper into the groove.

Water and counterions.

In the final model, the solvent region of DDD^{4+} contains 114 water molecules, eight partially occupied Tl^+ ions and two partially ordered spermine molecules. The $|F^+|$-$|F^-|$ map shows anomalous peaks from all eight of Tl^+ sites and from the DNA phosphorous atoms (Figure 21). The occupancies of the Tl^+ ions range from 0.15 to 0.22 with a summed occupancy of 1.35.

We have allocated the eight observed Tl^+ sites of DDD^{4+} to three classes: *conserved*, *semi-conserved*, and *variable* (Table 11 and 12). This classification scheme is based on similarities and differences in Tl^+ position of DDD^{4+} versus DDD. The Tl^+ sites of DDD were determined from crystals grown in the presence of thallium (I) acetate (Howerton et al., 2001). *Conserved* DDD^{4+} Tl^+ ions are less than 1Å from a DDD Tl^+ ion. *Semi-conserved* DDD^{4+} Tl^+ ions are between 1.0 and 4.0 Å from one or more DDD Tl^+ ions. Finally, *variable* DDD^{4+} Tl^+ ions are greater than 4.0 Å from any DDD Tl^+ ion.

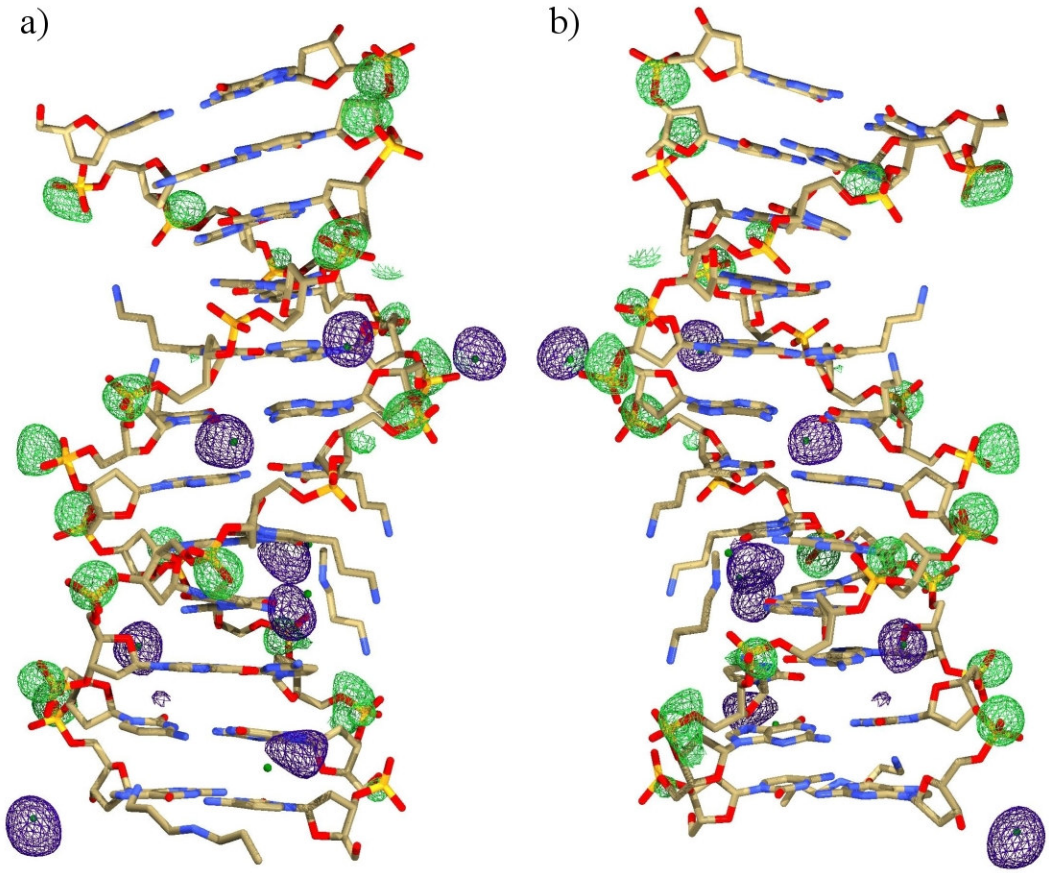

Figure 21. Anomalous density contoured at 2.5 σ around $^{XX}DDD^{4+}$.
The density around all atoms is drawn in green net, except for Tl^+ ions where it is drawn in blue net. Tl^+ ions are drawn as small green spheres. The figure shows the view of minor (a) and major (b) grooves.

Proximities were determined by superimposing DDD onto DDD^{4+}, as shown in Figure 19, then applying symmetry, then taking differences in positions. Only one Tl^+ site of DDD^{4+} falls in the *conserved* class. A Tl^+ ion is located in minor groove of both DDD^{4+} and DDD, at the ApT step. A Tl^+ ion with an occupancy of 0.15 in the DDD^{4+} makes contacts with four DNA oxygen atoms (two O2 atoms, and two O4' atoms) and two solvent sites. This hexacoordinate site is called the P3 site (Shui et al., 1998b; Woods et al., 2000), and is occupied by monovalent cations in a series of crystallographic studies of DDD, by a variety of cations (K^+, Rb^+, Cs^+, Tl^+) (Howerton et al., 2001; Shui et al., 1998b; Sines et al., 2000; Tereshko et al., 1999; Woods et al., 2000).

Table 11. Classification and location of $^{XX}DDD^{4+}$ Tl^+ Ions

Class[a]	Tl^+ in $^{xx}DDD^{4+}$	Tl^+ in DDD-Tl^+	Distance (Å)	Location of DDD^{4+} Tl^+
C	30	2106	0.22	Minor Groove, ApT-step
S	31	2104	3.04	Minor Groove
S	27	2105	3.50	Minor Groove
V	34	2101	4.80	Minor Groove
V	33	2110	4.91	Minor Groove
V	28	2113	5.02	
V	28	2110	5.57	Above Major Groove
V	28	2102	5.97	
V	29	-	>6	Minor Groove
V	32	-	>6	Edge of Major Groove

a) C: Conserved S: Semi-conserved V: Variable

Table 12. Positions and occupancies of DDD-Tl^+ and $^{xx}DDD^{4+}$ Tl^+ ions.

Molecule	$^{xx}DDD^{4+}$			DDD-Tl^+		
Region	Minor	Major	Other	Minor	Major	Other
Tl^+ Ion[a]	27(0.22)	33(0.15)	32(0.15)	2104(0.20)[d]	2101(0.34)	2111(0.18)
	29(0.20)	34(0.15)	28(0.15)	2105(0.18)[d]	2102(0.29)	2112(0.15)
	30(0.16)[c]			2106(0.10)[c]	2103(0.20)	
	31(0.16)			2109(0.12)[d]	2107(0.10)[d]	
					2108(0.10)[d]	
					2110(0.15)	
					2113(0.16)	
Total[b]	4(0.74)	2(0.30)	2(0.30)	4(0.60)	9(1.34)	2(0.33)

a) Tl^+ number is given with its occupancy in parentheses. b) Total number of Tl^+ ions are given with the sum of occupancies in parentheses. c) Located at the same site. d) Located near the bending region.

Of the remaining seven Tl^+ sites of DDD^{4+}, two are *semi-conserved*, and five are *variable*. In general, the sites in the minor groove are either *conserved* or *semi-conserved*, while the sites in the major groove are *variable*. The variability of these sites in the major groove is expected, because that is where the tethered cations are located. In fact, major groove of DDD^{4+} contains fewer cations than that of the native DDD. There are three Tl^+ ions in the major groove of the modified structure (28, 33, 34) of which, Tl 28 is well above the opening of the groove and should not be considered "inside" the groove. In comparison, there are seven Tl^+ ions in the major groove of DDD (2101, 2102, 2103, 2107, 2108, 2110, and 2114). It must be noted that Tl 2107 and 2108 are in

a region partially occupied by an Mg^{2+} coordinated with six water molecules. DDD^{4+} lacks an Mg^{2+} ion at this position. This is not necessarily an indication of altered electrostatic environment since DDD^{4+} was crystallized in the absence of Mg^{2+}. Whether this specific site can be occupied by an Mg^{2+} in DDD^{4+} structure must be determined through further experiments. However, the results obtained thus far do indicate that cations appear to have been expelled from the major groove, especially from the vicinity of the aminopropyl group of Z3dU 20. Some of the differences in Tl^{+} distribution between DDD^{4+} and DDD may arise from the methods used to introduce Tl+ ion into the crystals. The DDD^{4+} structure was soaked in Tl^{+}, while the DDD crystal was grown from a solution containing Tl^{+}.

Figure 19 shows a comparison of Tl^{+} distribution in the DDD^{4+} and DDD structures. Only the Tl^{+} at the P3 site in the minor groove is conserved in both structures. It appears to have similar occupancies, 0.15 in the modified and 0.10 in the unmodified structures. Table 12 contains the position and occupancy information for all of the Tl^{+} ions. There are three Tl^{+} ions in the major groove of the DDD structure, near the bending region (2103, 2107, and 2108). Tl 2107 and 2108 are at the same site that is partially occupied by an Mg^{2+} ion. In the DDD^{4+} structure, there is only one Tl+ ion (32) nearby, which is well above the major groove at a distance of 4.96 Å from $20N^{+}$. At the other terminus, there are four Tl+ ions in the DDD structure (2101, 2102, 2110, and 2113) and two in the DDD^{4+} (33 and 34). Tl 2102 and 2113 sites overlap and are mostly likely the same cation in two different positions. Tl 33 and 34 do not contact the floor of the major groove. Tl 33 is 6.24 Å from $7N^{+}$ and Tl 34 is 5.54 Å from $8N^{+}$. Another effect of this cation redistribution can be seen in the occupancies (Table 12). The total monovalent

cation occupancy in the minor groove is reasonably similar between the two structures (DDD^{4+} 0.74, DDD 0.60). However, the total monovalent cation occupancy is reduced in the major groove (DDD^{4+} 0.30, DDD 1.34). Interestingly, the total charge in the distorted region of the major groove is constant between the two structures (+2.0). This is considering the partially occupied Mg^{2+} ion and the two aminopropyl modifications. However, in DDD the charged atoms are distributed on bottom of the major groove and are in contact with the floor of the major groove, while in DDD^{4+} the charged species are well above the floor and distributed along the upper region of the major groove.

The final model of DDD^{4+} contains two partial spermine molecules labeled spermine 26 and spermine 27. The portions of these molecules that are most disordered are not visible in the electron density maps and have been left out of the model. The $2|F_o|-|F_c|$ electron density surrounding the spermine molecules is shown in Figure 22. Models in which the spermine molecules are replaced with water molecules show worse statistics and inferior maps. Refinement of a model omitting the spermine molecules resulted in strong (greater than 3.5 σ) difference $|F_o|-|F_c|$ density at their location. Both spermine molecules are located in the G-tract major groove on the end of the duplex where the conformation of DDD^{4+} is similar to that of native DDD. Spermine 25 stretches at the mouth of the major groove, making contacts with an adjacent DNA molecule. Spermine 26 extends outward from the major groove and is aligned with the aminopropyl modifications or7 and 8. In the cavity occupied by this spermine molecule and the two aminopropyl groups, there are ordered water molecules in the region proximal to the modifications, but none can be identified in the distal region of the cavity.

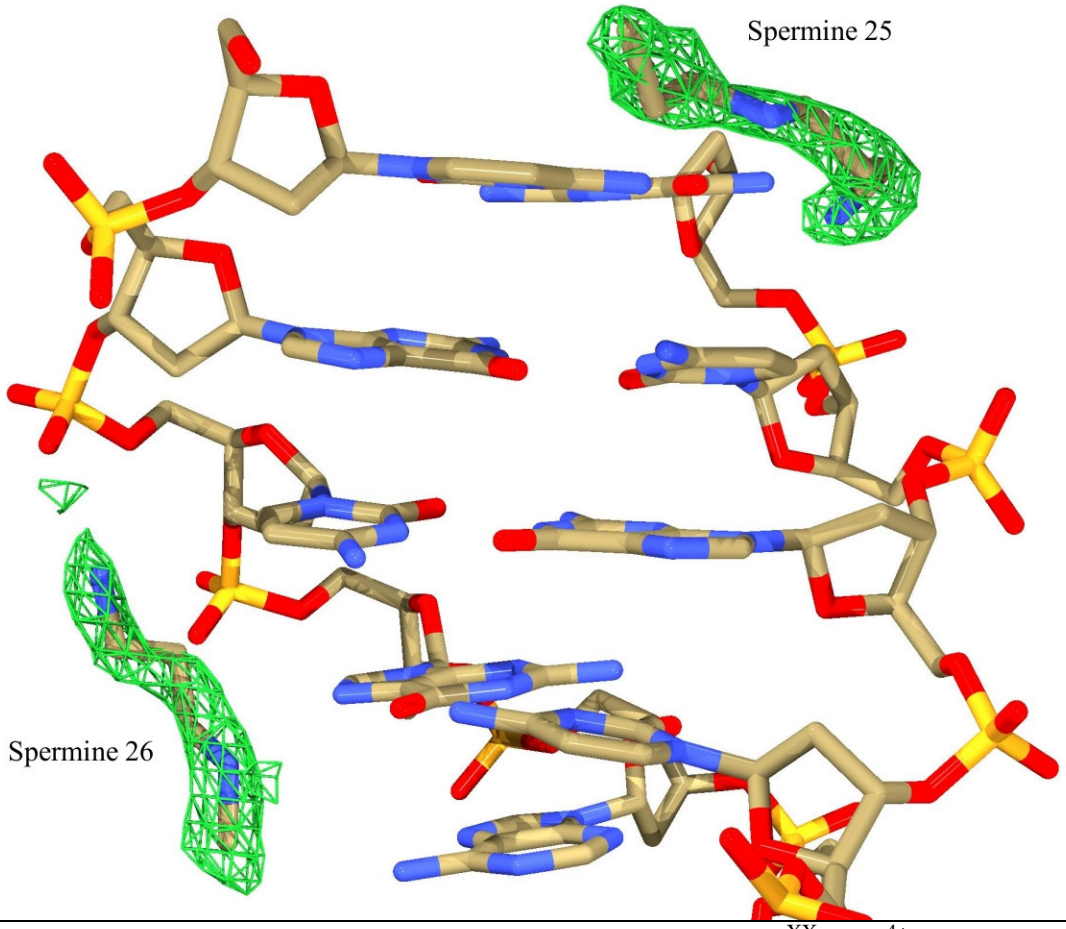

Figure 22. Sum electron density around spermine 25 and 26 in $^{XX}DDD^{4+}$.
Sum electron density is contoured at 1.0 σ (green net) around the two partial spermine molecules.

It is likely that disordered solvent molecules, and perhaps the remainder of spermine 26 occupy the rest of this cavity.

There are two components in the crystallization solution with the topography of an unbranched chain, spermine and the detergent ANAPOE™-$C_{12}E_{10}$ (polyoxyethylene(10)dodecyl ether). As partial molecules, they may be difficult to distinguish. The electron density associated with the spermine molecules might also be modeled as partial detergent molecules. However the detergent molecules are 42 atoms long, and more hydrophobic than spermine. Spermine appears to be the more logical

choice for the small, hydrophilic environment of the inter-molecular cavities of the crystal.

Discussion

Double helical DNA is a semi rigid polymer with a persistence length of 140-150 bp (Hagerman & Hagerman, 1996; Hagerman, 1988). Under certain conditions, DNA can be induced to bend spontaneously. DNA bends away from TATA-element binding protein (TBP) by almost 90° (Juo et al., 1996; Nikolov et al., 1996). DNA bends toward catabolite activator protein (CAP) by around 90° (Schultz, Shields & Steitz, 1991). In the nucleosome core particle DNA wraps around a complex of histone proteins by nearly two full turns (Kornberg & Klug, 1981; Luger et al., 1997; Van Holde, Sahasrabuddhe & Shaw, 1974).

Many of these DNA distortions can be explained by the electrostatic collapse model of Mirzabeckov, Rich and Manning. In one test of that model 'phantom proteins' were created by Maher (Strauss & Maher, 1994; Strauss et al., 1996a; Strauss-Soukup & Maher, 1997; Strauss-Soukup & Maher, 1998; Strauss-Soukup, Rodrigues & Maher, 1998). DNA was covalently tethered to cations – i.e., phantom proteins. Maher also partially neutralized DNA by modification of phosphates. From the effects of the phantom proteins on the electrophoretic mobility of DNA, it was concluded that axial bending can be caused by interactions of charged species with DNA. Appropriate cation adduction or phosphate neutralization induces spontaneous bending of DNA, suggesting the importance of asymmetric charge neutralization.

The effects of tethered cations on three-dimensional structure of DNA at high resolution have been studied by Li et al., who obtained NMR results using the singly modified d(CGCGAATXCGCG)]$_2$ ($^{TX}DDD^{2+}$) molecule. The results are consistent with DNA deformation, including axial bending (Li et al., 2002). The NMR data suggest an ordered environment for the Z3dU moiety within the major groove. The amino propyl group appears to extend in the 3' direction, accompanied by a perturbation of the phosphodiester backbone of the DDD^{2+} molecule. This perturbation is indicated by a chemical shift perturbation of the ^{31}P resonance of the relevant phosphodiester linkage. The DNA deformations were attributed to an electrostatic interaction between phosphates and the tethered cationic charge.

In the DDD^{4+} structure described here, four positively charged amino groups are tethered to the major groove of each dodecamer duplex. Three of the tethered cations are directed radially out from the DNA, and do not appear to interact significantly with the DNA. These tethered cations do not appear to induce structural changes or to displace counterions. On one end of the duplex, spermine molecules and Tl^+ ions are located on the major groove floor, in regions similar to those in the DDD control, the tethered amino groups are remote from the DNA. The DNA conformation is unperturbed.

By contrast on the opposing end of the duplex, one of the tethered cations is directed in the 3' direction toward a phosphate group. This tethered cation appears to interact electrostatically with the DNA. This interaction is accompanied by changes in helical parameters rise, roll and twist, and by a displacement of the backbone relative to a control oligonucleotide. In addition these interactions appear to be associated with displacement of counterions from the major groove of the DNA.

In the NMR solution structure of the DDD^{2+}, the DNA shows two-fold symmetry, as anticipated. The NMR structure represents a time-average of the solution states of the molecule. The NMR experiment cannot distinguish between the two termini. The two aminopropyl modifications behave similarly and interact with the backbone to bend the molecule at both termini. The x-ray structure of DDD^{4+} lacks two-fold symmetry. The end of the duplex with greatest cation-phosphate proximity shows alteration in structure, relative to the control DDD. Lattice forces in the crystal state create two different electrostatic environments at the termini. It is possible the DDD^{4+} structure may represent an intermediate in the solution bending process. In summary, the structural studies are consistent with the Rouzina and Bloomfield model for collapse of the major groove around bound cations. However, in is conceivable that the observed conformation is a 'trapped' intermediate. In solution the disortion may be greater than in the crystal structure described here. Further studies on how the sequence-dependent location of cations affects DNA structure are ongoing.

It was proposed that tethered cations ions would form salt bridges with the O1P or O2P of the phosphates. The term 'salt bridge' indicates a van der Waals contact between two atoms carrying full or partial positive and negative charges. However, the proposed conformation required for salt bridging appears was not confirmed by electrostatic footprinting, molecular modeling and NMR studies. Those results support conformations where the tethered cations are oriented in the 3'-direction. The orientation of the tethered ammonium ion positions the ammonium ion near the floor of the major groove, in proximity to a G, or into the solution in a sequence dependent manner (Dande et al., 1997; Heystek et al., 1998). This orientation appears to be confirmed here. However it

should be noted that electrostatic interactions fall off very gradually with distance. An absence of salt bridges does not imply absence of strong electrostatic interaction.

CHAPTER VI

COMPARISON OF TWO B-DNA STRUCTURES WITH TETHERED CATIONS

Abstract

This chapter is based on a manuscript prepared for publication. This work was performed in collaboration with Sarah Tannenbaum, Tatsuya Maehigashi, Dr. Barry Gold, Dr. Micheal Stone, Dr. Luis Marky and Dr. Loren Williams. Tatsuya Maehigashi kindly collected the data for this work. The X-ray crystal structure of the dodecamer $[d(CGCGAATXCGCG)]_2$ (X = Z3dU, 5-(3-aminopropyl)-2'-deoxyuridine, Figure 3, Chapter II) was determined at pH of 6.7. At this pH, the amino group of the modification is protonated and carries a charge of +1. The structure shows that the tethered amino groups extend into the intermolecular cavities within the crystal and do not make significant contacts with the backbone. This is in contrast to the NMR structure of the same dodecamer and the x-ray crystal structure of a similar dodecamer, $[d(CGCGAAXXCGCG)]_2$. The evidence from these structures suggests that the backbone is deformed. The x-ray crystal structure does not demonstrate any significant deviation from the canonical B-DNA conformation. The inconsistency between the crystal and NMR structures is most likely due to packing effects.

Introduction

There are many factors that determine the DNA conformation. In a ground state structure (i) interactions of water molecules with the hydrophobic planer surfaces of

DNA bases are minimized, (ii) stacking and hydrogen-bonding interactions are maximized, and (iii) intra-molecular phosphate-phosphate repulsions are minimized. It is empirically known that DNA sequence is an important modulator of conformation. A-tracts are associated with narrow minor grooves (Alexeev et al., 1987; Burkhoff & Tullius, 1987; Wing et al., 1980), axial bends (Diekmann & Wang, 1985; Hagerman, 1984; Marini et al., 1982), and high propeller twists (Drew et al., 1981). G-tracts exhibit strong divalent cation dependence of bending (Brukner et al., 1994) and a propensity to convert to A-form (Hud & Plavec, 2003).

The sequence-dependence of DNA conformation can be ascribed to base-specific hydrophobicity, shape, charge distribution, and functional group disposition. The relative importance and mechanisms by which various sequence-specific factors influence DNA conformation is not currently resolved. The goal of the work here is to determine the effects of tethering cationic charge to DNA on conformation, and on distribution of non-tethered counterions such as K^+, Mg^{2+} and spermine.

We have determined the x-ray structure of the DNA duplex CGCGAATXCGCG, ($^{TX}DDD^{2+}$, where 2^+ indicate total charge of tethered cations and not the total charge of the DNA molecule, -18). We attempted to crystallize the duplex CGCGAAXTCGCG ($^{XT}DDD^{2+}$). For ease of communication, these two dodecamers will be distinguished by the appropriate superscript (TX or XT). Unlike the solution NMR structure of $^{TX}DDD^{2+}$ and the x-ray structure of $^{XX}DDD^{4+}$, we see no axial bending and observe no evidence of interaction between the ω-amino group of the tether and either the backbone or the floor of the major groove.

We attribute this discrepancy to crystal-packing effects, which prevent DNA bending. By comparing the position of symmetry related molecules in crystals of $^{TX}DDD^{2+}$ and $^{XX}DDD^{4+}$, we demonstrate that the orientation of $^{XX}DDD^{4+}$ in the unit cell allows for collapse of the backbone. However, the orientation of $^{TX}DDD^{2+}$ 'locks' the backbone in a straight conformation. Li and co-workers have reported evidence for the interaction of the amino group of the Z3dU tether with the backbone (Li et al., 2002). In contrast to their findings and the crystal structure of $^{XX}DDD^{4+}$, the aminopropyl moieties tethered to the major groove of $^{TX}DDD^{2+}$ extend into the intermolecular cavities of the crystal. They are oriented in the 3'-direction and are disordered.

The motivation for this work stems ultimately from work of Mirzabeckov, Rich and Manning (Manning et al., 1989; Mirzabekov & Rich, 1979), who considered the proposition that charge neutralization of phosphate groups might drive DNA bending. They were concerned with asymmetric charge neutralization by cationic protein sidechains, leading to 'phosphate collapse'. This model was extended by Rouzina and Bloomfield (Rouzina & Bloomfield, 1998), who proposed that mobile cations such as Mg^{2+} can contribute electrostatically to DNA bending. In the Rouzina and Bloomfield model, a divalent cation can repel and displace other counterions, leading to strong attraction with unscreened phosphate groups. The result is DNA bending by electrostatic collapse around a divalent cation. Our experimental system has limitations for analysis of DNA bending. However, it does provided a detailed view of microheterogenity in DNA conformation.

Materials and Methods

Crystallization.

The 12-mer d(CGCGAATXCGCG) was synthesized and purified as described (Heystek et al., 1998; Li et al., 2002). The oligomer was annealed by heating to 80ºC and slow cooling to room temperature. Concentration of the annealed DNA was determined by UV spectroscopy. Crystals were grown over 2-3 days using the sitting drop vapor diffusion method. The crystallization solution contained 1.2 mM DNA duplex, 8.0 mM magnesium acetate (pH 6.5), 4.0% v/v 2-methyl-2,4-pentanediol (MPD), and 16.0 mM spermine acetate (pH 6.5). The crystallization solution was equilibrated against a reservoir of 35% v/v MPD at 22ºC. A crystal of dimensions 0.2 x 0.2 x 0.1 mm^3 was looped and frozen in liquid nitrogen.

Data collection and processing.

180º of data, with oscillation angle of 1.0º was collected at beamline 22-ID at the SER-CAT facilities at APS using a MAR-165 CCD detector at a wavelength of 1.00931 Å. The crystal was maintained at -160ºC. The data were scaled and integrated using HKL2000 version 1.97.7 (Otwinowski & Minor, 1997).

Inspection of the images showed that the crystal was twinned. Data obtained by integrating all frames generated poor maps with missing sum electron density throughout the structure (Figure 4, Chapter II). The quality of the data was significantly improved by excluding a subset of frames from scaling based on the statistics that HKL2000 produced on detector-to-crystal distance, unit cell geometry, and crystal rotation. Frames 76-80, 96-105, and 116-120 of the 180 frames were excluded from the scaling process. An initial round of integration was performed using all frames. In the final round of

integration, the unit cell parameters obtained from the initial integration round were fixed, and the twinned frames were excluded. A total of 115580 reflections were indexed, integrated, and reduced to 10299 unique reflections. Unit cell dimensions are a = 25.29 Å, b = 40.10 Å, and c = 65.62 Å, in space group $P2_12_12_1$. The data used in the refinement included 9233 unique reflections from 75 to 1.51 Å.

Refinement.

Molecular replacement was used for phase determination. The coordinates for the starting DNA model were obtained from NDB (entry BDL084). The model was annealed and refined against the data using the program CNS (version 1.1) and parameters of Berman and coworkers (Clowney et al., 1996; Gelbin et al., 1996; Parkinson et al., 1996).

The parameters were modified to remove all existing dihydral angle restraints for the backbone and to add new dihydral restraints to keep the purine and pyrimidine rings planar. Restraints for Z3dU were estimated by combining elements of thymine and lysine. After one cycle of refinement, the atomic coordinates for the three terminal atoms in the aminopropyl modification of Z3dU8 could be deduced from the difference electron density maps. T7 of the original model was converted to a Z3dU residue incorporating these coordinates. Water molecules were added iteratively followed by annealing, refinement, and phase calculation using peaks in the sum and electron density maps. The position of the aminopropyl modification of Z3dU20 was more difficult to discern. The atomic coordinates for this modification were incorporated into the model after the majority of the water molecules were assigned.

Table 13. Crystallographic and refinement statistics for $^{TX}DDD^{2+}$	
unit cell	
a (Å)	25.29
b (Å)	40.10
c (Å)	65.26
space group	$P2_12_12_1$
temp. of data collection (°C)	-160.0
no. of reflections	115580
no. of unique reflections	10299
completeness (%)/highest shell (%)	93.5/71.7
max resolution of obsd reflections (Å)	1.49
max resolution of highest shell used (Å)	1.51
resolution range used in refinement (Å)	35-1.51
no. of reflections used in refinement	9233
no. of reflections used in the test set	1066
rmsd of bonds from ideal (Å)	0.0098
rmsd of angles from ideal (Å)	1.4439
DNA (asymmetric unit)	$[d(CGCGAATXCGCG)]_2$
no. of DNA atoms	492
no. of water molecules, excluding Mg first shell	162
no. of Mg ions plus coordinating first shell	7
no. of spermine atoms	0
R-free (%)	22.79
R-factor (%) excluding test set data	18.92
unit cell volume occupied (%)	68.59

Results

The final model contains 162 water molecules and one $Mg^{2+}(H_2O)_6$ (Figure 23). The magnesium ion is bound in the major groove as observed in other low magnesium concentration structures of DDD (Shui et al., 1998a; Shui et al., 1998b; Woods et al., 2000). The statistics for data refinement (Table 13) and the quality of the sum electron

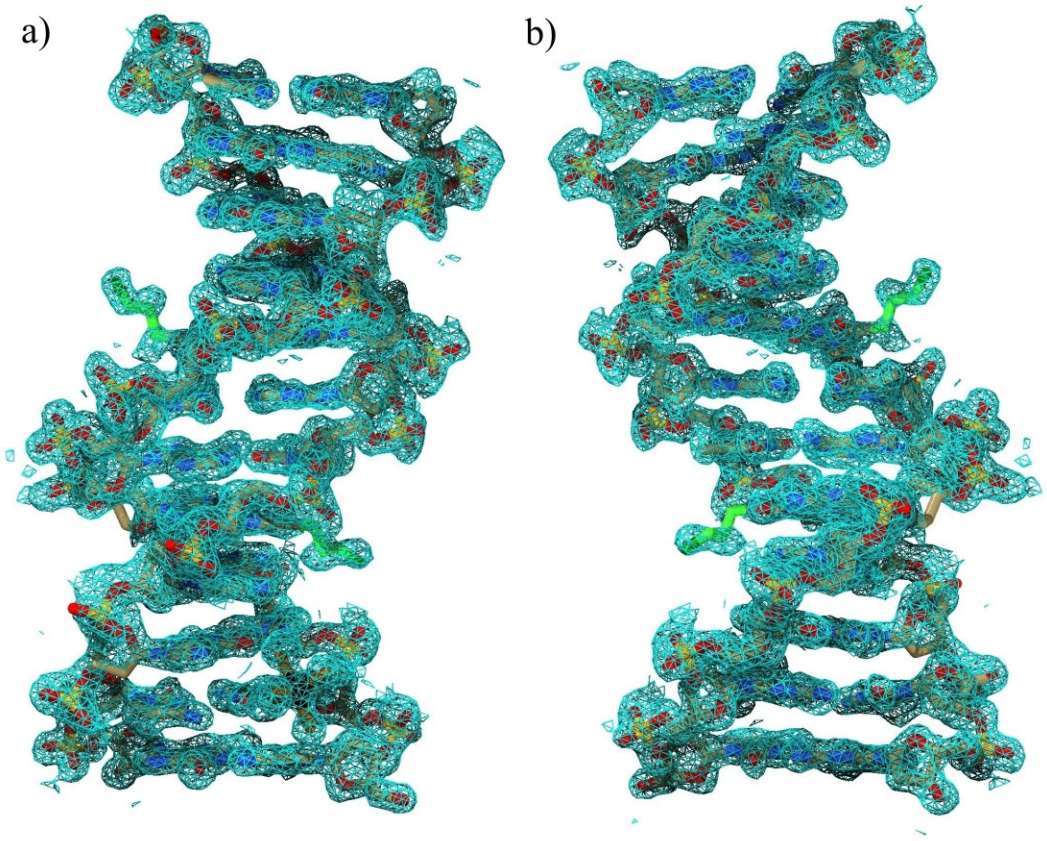

Figure 23. Sum electron density contoured at 1.0 σ around $^{TX}DDD^{2+}$.
The sum electron density (cyan net) is contoured at 1.0 σ. The aminopropyl modifications of Z3dU8 and Z3dU20 are drawn in green with the N5D in dark green. The figure shows the view of minor (a) and major (b) grooves.

density map (Figure 23) show that the model is in agreement with the data. However, not all atoms of the tethered cations of Z3dU's are within electron density (Figure 24). Both modifications are disordered. Attempts were made to refine the model with multiple conformations of the Z3dU modification. These refinements did not improve the electron density maps or the refinement statistics. Therefore, the final position of the two modifications is their major conformation as deduced from the electron density maps. There is no deformation of the backbone as observed in the NMR solution structure of $^{TX}DDD^{2+}$ (Li et al., 2002) and the x-ray structure of $^{XX}DDD^{4+}$ (Figure 25).

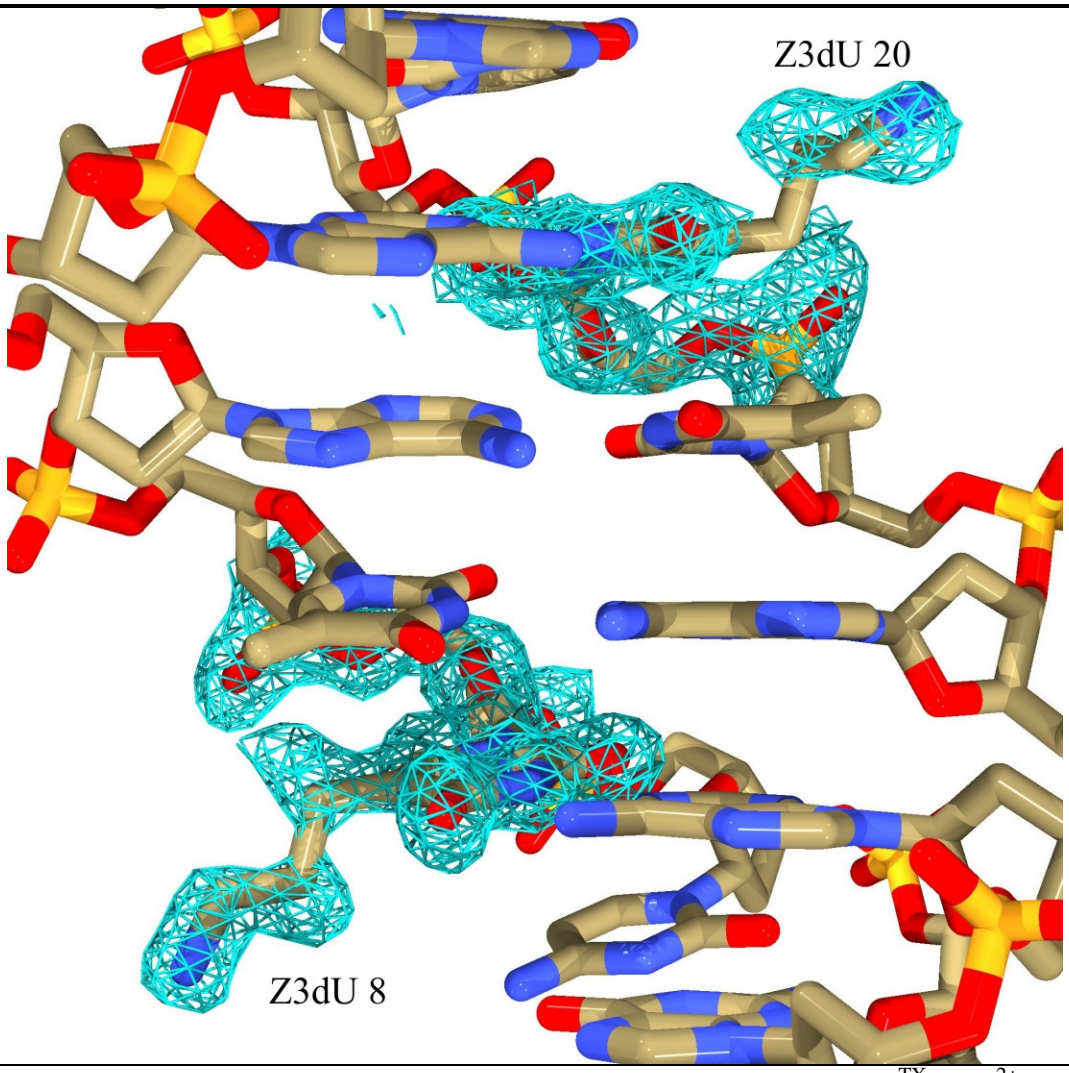

Figure 24. Detailed view of sum electron density around the 2 Z3dU in $^{TX}DDD^{2+}$.
Sum electron density is contoured at 1.0 σ and drawn in cyan net around Z3dU 8 and 20.

Furthermore, the Z3dU8 modifications in $^{TX}DDD^{2+}$ and $^{XX}DDD^{4+}$ are generally in the same position. In contrast, the Z3dU20 modifications in $^{TX}DDD^{2+}$ and $^{XX}DDD^{4+}$ are in slightly different orientation. The aminopropyl group of Z3dU20 in $^{TX}DDD^{2+}$ has the same orientation seen for the modifications of all Z3dU20 residues, which do not interact with the backbone, directed into the intermolecular cavities. These non-interacting tethered cations can be superimposed and show small variations in position, but in

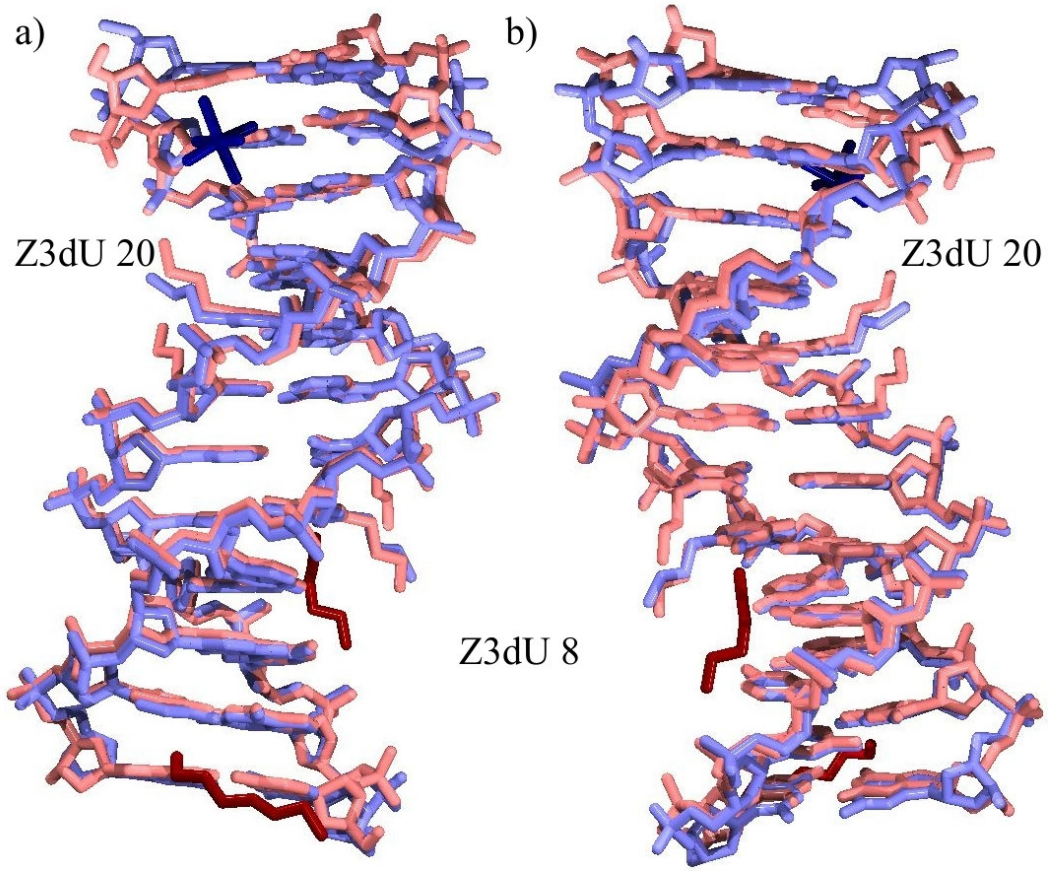

Figure 25. Superimposed structures of $^{TX}DDD^{2+}$ and $^{XX}DDD^{4+}$.
The structures of $^{TX}DDD^{2+}$ and $^{XX}DDD^{4+}$ are superimposed showing the view of minor (a) and major (b) grooves. $^{TX}DDD^{2+}$ is in light blue with its $Mg^{2+}(H_2O)_6$ ion in dark blue. $^{XX}DDD^{4+}$ is in pink with its spermine molecules in red.

general have the same orientation with respect to the plane of their covalently linked Z3dU base However, the aminopropyl group of Z3dU20 in $^{XX}DDD^{4+}$ has collapsed towads the opposite backbone and appears to be positioned further inside the major groove.

It merits mention that in $^{TX}DDD^{2+}$, the Z3dU20 tethered cation is more disordered than the Z3dU8 tethered cation (Figure 24), since (i) the sum electron density around Z3dU20 modification is not as intense and connected as the density around Z3dU8 modification, and (ii) there are more alternative conformations plausible for the

modification of Z3dU20 than for the Z3dU8 modification (as deduced from the electron density maps). The tethered cation of Z3dU8 is in a narrower cavity with well-ordered water molecules in its immediate surrounding. The sum electron density assigned to this modification suggests that it exists in two conformations. In the dominant conformation (as it is modeled), the modification is extended in the 3' direction. In the second conformation (not modeled), it appears that the C5-C5A bond is rotated by 180°, orienting the modification in the 5' direction. The Z3dU20 modification is more disordered, with no density observed around the C5A carbon.

We believe that the lack of observed backbone deformation in x-ray compared to NMR structure of $^{TX}DDD^{2+}$ is due to lattice packing effects. In order to explore this hypothesis, we compared the symmetry related molecules of $^{TX}DDD^{2+}$ and $^{XX}DDD^{4+}$ near the bending region in $^{XX}DDD^{4+}$. To do this $^{TX}DDD^{2+}$ and $^{XX}DDD^{4+}$ were superimposed using only the central A-tract region of the two molecules and then symmetry was applied using each molecules corresponding unit cell parameters.

Figure 26 shows $^{TX}DDD^{2+}$ (blue), $^{XX}DDD^{4+}$ (red), a duplex related by symmetry to $^{TX}DDD^{2+}$ (cyan), and duplex related by symmetry to $^{XX}DDD^{4+}$ (blue). The inset shows the detail of the interactions. In the $^{TX}DDD^{2+}$ crystals, a symmetry related molecule (cyan) stericly prevents further movement of $^{TX}DDD^{2+}$ backbone (blue) into the major groove. This steric interference is reduced but not eliminated in the crystals of $^{XX}DDD^{4+}$. The $^{XX}DDD^{4+}$ backbone (red) has more space to move into the backbone, because its symmetry related duplex (magenta) is in a different orientation.

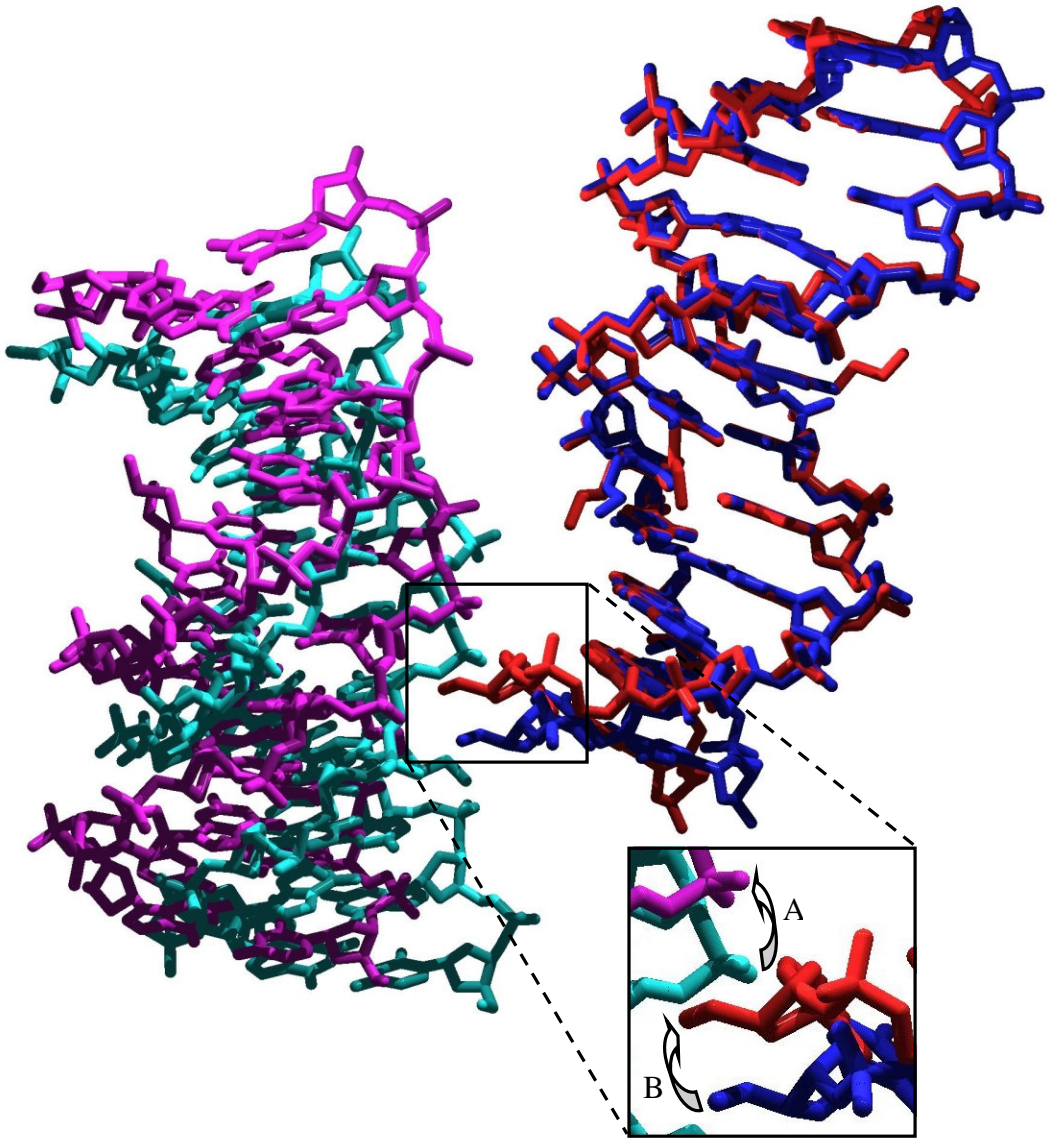

Figure 26. Comparison of the crystal packing of $^{XX}DDD^{4+}$ and $^{TX}DDD^{2+}$.
$^{TX}DDD^{2+}$ (blue) is superimposed on $^{XX}DDD^{4+}$ (red) and their symmetry related molecules are generated. The duplex related to $^{TX}DDD^{2+}$ by symmetry is in cyan. The duplex related to $^{XX}DDD^{4+}$ is in magenta. The inset shows the details of the steric interference. A change in the orientation of the symmetry related molecules (arrow A) allows for a shift in the position of the backbones (arrow B)

Discussion

The most likely cause for the straight x-ray structure of $^{TX}DDD^{2+}$ as compared to the NMR structure of $^{TX}DDD^{2+}$ is the crystal lattice packing effects. These effects may 'lock' the DNA molecules into one conformation. This possibility is illustrated in Figure

26. The lack of deformation of the backbone in $^{TX}DDD^{2+}$ is due to a steric interference from a symmetry related molecule. This interference is reduced in the crystals of $^{XX}DDD^{4+}$, because of reorientation of the duplex molecules. This reorientation (arrow A in Figure 26) creates more space available to the backbone of $^{XX}DDD^{4+}$ and as a result the phosphodiester backbone of G2 can move further into the major groove (arrow B). In $^{TX}DDD^{2+}$, the symmetry related molecule is positioned such that the backbone is prevented from any further movement towards the inside of the major groove (blue and cyan molecules in Figure 26).

Another unexpected result is that there is sum electron density around all atoms of the modification on Z3dU8, but only around the three terminal atoms of the modification on Z3dU20. Looking at the environment of the two modifications hints at one possible explanation. The aminopropyl moiety of Z3dU8 is located in a narrower cavity than that of Z3dU 20. The solvent molecules around Z3dU8 are more ordered, restricting movement of the tethered group. In case of Z3dU20, the cavity is wider with less ordered solvent molecules, allowing the aminopropyl moiety to adapt more conformations.

Still, it's surprising that there is no sum density around C5A of Z3dU20, which lies in the plane of the prymidine base. The absence of any density around this atom may be due to errors in the Fourier synthesis process or due to the quality of the data.

APPENDIX I

DATA PROCESSING STATISTICS GENERATED BY HKL2000

Statistics for $^{XX}DDD^{4+}$ Structure (Crystal tm60B1a)

```
number of film packs  180
 total number of reflections used   146309
 Goniostat axis directions:
   phi   0.00000   0.00000  -1.00000
 kappa  -1.00000   0.00000   0.00000
 omega   0.00000   1.00000   0.00000
 Goniostat angles:
  o-detector,    o-beam,    o-zero,   o-kappa,    k-zero, phi-kappa
       0.000      0.000    -90.000     90.000    -90.000     90.000
 Space group p212121 number  19 number of symmetry operators  4
 x, y, z;
 1/2-x, -y, z+1/2;
 x+1/2, 1/2-y, -z;
 -x, y+1/2, 1/2-z;
Unit cell volume     68076.9

 New scale
     1   0.9594     2   0.9658     3   0.9777     4   0.9644     5   0.9848
     6   0.9828     7   1.0000     8   1.0041     9   1.0099    10   1.0285
    11   1.0024    12   1.0190    13   1.0229    14   1.0570    15   1.0612
    16   1.0710    17   1.0809    18   1.0909    19   1.1010    20   1.1112
    21   1.1216    22   1.1029    23   1.0999    24   1.1291    25   1.1540
    26   1.1355    27   1.1206    28   1.1690    29   1.1437    30   1.1641
    31   1.1874    32   1.1699    33   1.2179    34   1.2195    35   1.1980
    36   1.1998    37   1.2297    38   1.2231    39   1.2197    40   1.2165
    41   1.2300    42   1.2359    43   1.2201    44   1.2295    45   1.2359
    46   1.2570    47   1.2506    48   1.2276    49   1.2333    50   1.2326
    51   1.2149    52   1.2360    53   1.2356    54   1.2383    55   1.2258
    56   1.2096    57   1.2185    58   1.2060    59   1.2152    60   1.2059
    61   1.2250    62   1.2165    63   1.1969    64   1.2021    65   1.1920
    66   1.2199    67   1.1911    68   1.1783    69   1.2064    70   1.2007
    71   1.1911    72   1.1846    73   1.1944    74   1.1715    75   1.1873
    76   1.1918    77   1.1780    78   1.1624    79   1.1700    80   1.1691
    81   1.1693    82   1.1565    83   1.1558    84   1.1582    85   1.1714
    86   1.1558    87   1.1431    88   1.1653    89   1.1458    90   1.1479
    91   1.1573    92   1.1263    93   1.1303    94   1.1480    95   1.1232
    96   1.1311    97   1.1412    98   1.1437    99   1.1373   100   1.1342
   101   1.1240   102   1.1338   103   1.1300   104   1.1135   105   1.1218
   106   1.1108   107   1.1073   108   1.1146   109   1.1282   110   1.1049
   111   1.1091   112   1.1041   113   1.1014   114   1.0993   115   1.0978
   116   1.0978   117   1.1106   118   1.0979   119   1.0826   120   1.0970
   121   1.1163   122   1.0988   123   1.1168   124   1.0989   125   1.0994
   126   1.1020   127   1.1046   128   1.1072   129   1.1098   130   1.1124
   131   1.1150   132   1.1253   133   1.1331   134   1.1321   135   1.1710
   136   1.1408   137   1.1402   138   1.1581   139   1.1419   140   1.1797
   141   1.1861   142   1.1780   143   1.1933   144   1.1944   145   1.1901
   146   1.2225   147   1.2084   148   1.2144   149   1.2217   150   1.2538
   151   1.2388   152   1.2398   153   1.2562   154   1.2456   155   1.2545
   156   1.2880   157   1.2805   158   1.2897   159   1.2754   160   1.2887
   161   1.3208   162   1.3142   163   1.3232   164   1.3380   165   1.3532
   166   1.3596   167   1.3486   168   1.3641   169   1.3912   170   1.3769
   171   1.3815   172   1.3742   173   1.4110   174   1.3883   175   1.4147
   176   1.3870   177   1.3982   178   1.4272   179   1.4415   180   1.4508

 Errors of the scale factors: film #, % error
     1      0.0     2      0.0     3      0.0     4      0.0     5      0.0
     6      0.0     8      0.0     9      0.0    10      0.0    11      0.0
    12      0.0    13      0.0    14      0.0    15      0.0    16      0.0
    17      0.0    18      0.0    19      0.0    20      0.0    21      0.0
    22      0.0    23      0.0    24      0.0    25      0.0    26      0.0
    27      0.0    28      0.0    29      0.0    30      0.0    31      0.0
    32      0.0    33      0.0    34      0.0    35      0.0    36      0.0
    37      0.0    38      0.0    39      0.0    40      0.0    41      0.0
```

42	0.0	43	0.0	44	0.0	45	0.0	46	0.0
47	0.0	48	0.0	49	0.0	50	0.0	51	0.0
52	0.0	53	0.0	54	0.0	55	0.0	56	0.0
57	0.0	58	0.0	59	0.0	60	0.0	61	0.0
62	0.0	63	0.0	64	0.0	65	0.0	66	0.0
67	0.0	68	0.0	69	0.0	70	0.0	71	0.0
72	0.0	73	0.0	74	0.0	75	0.0	76	0.0
77	0.0	78	0.0	79	0.0	80	0.0	81	0.0
82	0.0	83	0.0	84	0.0	85	0.0	86	0.0
87	0.0	88	0.0	89	0.0	90	0.0	91	0.0
92	0.0	93	0.0	94	0.0	95	0.0	96	0.0
97	0.0	98	0.0	99	0.0	100	0.0	101	0.0
102	0.0	103	0.0	104	0.0	105	0.0	106	0.0
107	0.0	108	0.0	109	0.0	110	0.0	111	0.0
112	0.0	113	0.0	114	0.0	115	0.0	116	0.0
117	0.0	118	0.0	119	0.0	120	0.0	121	0.0
122	0.0	123	0.0	124	0.0	125	0.0	126	0.0
127	0.0	128	0.0	129	0.0	130	0.0	131	0.0
132	0.0	133	0.0	134	0.0	135	0.0	136	0.0
137	0.0	138	0.0	139	0.0	140	0.0	141	0.0
142	0.0	143	0.0	144	0.0	145	0.0	146	0.0
147	0.0	148	0.0	149	0.0	150	0.0	151	0.0
152	0.0	153	0.0	154	0.0	155	0.0	156	0.0
157	0.0	158	0.0	159	0.0	160	0.0	161	0.0
162	0.0	163	0.0	164	0.0	165	0.0	166	0.0
167	0.0	168	0.0	169	0.0	170	0.0	171	0.0
172	0.0	173	0.0	174	0.0	175	0.0	176	0.0
177	0.0	178	0.0	179	0.0	180	0.0		

Summary of reflection intensities and R-factors by batch number

		All data			Linear
Batch	# obs	# obs > 1	<I/sigma>	N. Chi**2	R-fac
1	252	251	12.9	1.589	0.068
2	385	385	10.3	1.137	0.065
3	268	267	12.0	1.711	0.072
4	287	285	10.4	1.417	0.053
5	296	294	13.9	1.788	0.065
6	327	327	11.9	3.098	0.086
7	341	341	11.2	2.383	0.088
8	290	289	11.6	2.409	0.087
9	269	268	12.0	2.300	0.079
10	267	267	11.6	3.155	0.112
11	286	285	11.2	1.766	0.077
12	340	338	11.8	1.151	0.060
13	306	300	11.9	1.846	0.076
14	380	379	13.4	1.508	0.062
15	222	222	13.2	1.902	0.091
16	0	0	0.0	0.000	0.000
17	0	0	0.0	0.000	0.000
18	0	0	0.0	0.000	0.000
19	0	0	0.0	0.000	0.000
20	0	0	0.0	0.000	0.000
21	185	185	11.9	1.510	0.060
22	326	325	12.0	1.154	0.043
23	294	290	10.8	1.000	0.048
24	342	342	12.0	1.351	0.057
25	314	312	13.4	1.057	0.051
26	334	331	10.9	0.905	0.057
27	331	330	11.5	0.888	0.047
28	325	323	10.8	0.980	0.056
29	343	340	11.0	1.170	0.057
30	331	331	10.7	1.942	0.088
31	333	331	10.5	0.912	0.061
32	357	352	10.9	0.695	0.047
33	341	341	10.4	0.930	0.055
34	348	343	11.2	1.024	0.056
35	338	335	11.7	0.915	0.051
36	301	301	13.4	1.306	0.045
37	312	311	13.0	0.852	0.042
38	328	326	12.5	1.122	0.052
39	313	307	11.8	1.120	0.049
40	327	325	11.0	1.162	0.067
41	290	287	12.5	0.729	0.043
42	314	310	10.8	0.888	0.039
43	335	326	11.4	0.710	0.044
44	299	297	10.6	0.716	0.044
45	280	279	10.4	0.783	0.043
46	308	306	10.8	0.744	0.044
47	360	350	12.2	0.780	0.036
48	346	344	12.8	0.943	0.047
49	334	334	12.5	1.073	0.048

50	335	334	10.7	0.771	0.043
51	315	315	11.6	0.902	0.046
52	265	264	11.0	0.808	0.045
53	271	270	11.6	0.726	0.043
54	378	378	11.8	0.910	0.052
55	355	354	12.4	0.777	0.043
56	302	300	13.1	0.982	0.050
57	327	325	11.9	0.849	0.046
58	266	260	13.0	0.947	0.050
59	358	354	11.5	1.201	0.058
60	352	349	11.1	0.788	0.036
61	304	291	10.5	0.817	0.050
62	339	330	11.2	0.661	0.036
63	275	274	13.0	1.118	0.051
64	349	344	12.1	0.791	0.039
65	357	354	12.2	0.709	0.038
66	306	301	11.3	0.756	0.043
67	336	329	13.1	0.899	0.044
68	332	329	10.8	0.699	0.046
69	355	349	9.9	0.649	0.041
70	312	311	11.2	0.836	0.043
71	312	309	11.9	0.709	0.037
72	313	310	12.1	0.728	0.036
73	333	330	12.4	0.924	0.045
74	309	307	10.7	0.810	0.050
75	307	306	12.0	0.971	0.055
76	290	288	11.6	0.828	0.037
77	308	303	13.2	0.753	0.042
78	326	319	11.4	0.811	0.055
79	326	325	14.2	0.762	0.032
80	290	287	11.8	0.664	0.037
81	302	302	11.7	0.634	0.034
82	267	263	11.4	0.632	0.043
83	298	296	10.8	0.746	0.050
84	322	321	14.0	0.636	0.032
85	260	260	12.3	0.688	0.034
86	298	298	11.6	0.746	0.035
87	305	304	13.4	0.921	0.043
88	291	289	12.5	0.690	0.042
89	292	291	11.3	0.599	0.042
90	325	324	11.8	0.680	0.038
91	274	267	11.9	0.776	0.040
92	312	302	12.1	0.826	0.040
93	329	318	10.7	0.750	0.048
94	288	277	11.9	0.887	0.051
95	286	277	12.9	0.648	0.032
96	294	293	10.7	0.623	0.038
97	287	287	12.2	0.847	0.047
98	290	288	13.8	0.871	0.038
99	297	297	13.5	0.544	0.028
100	305	304	12.0	0.703	0.034
101	288	283	12.8	0.755	0.034
102	261	258	11.1	0.728	0.035
103	283	280	11.6	0.826	0.047
104	330	329	13.3	0.657	0.035
105	290	290	13.5	1.014	0.054
106	291	290	13.0	0.652	0.036
107	268	267	10.1	0.797	0.042
108	294	294	12.3	0.564	0.033
109	281	279	10.5	0.803	0.048
110	275	274	12.7	0.581	0.030
111	322	321	13.3	0.683	0.031
112	281	279	13.1	0.949	0.043
113	262	260	12.6	0.799	0.046
114	317	312	12.1	0.747	0.044
115	286	283	12.7	0.777	0.039
116	264	261	13.5	0.557	0.028
117	295	294	10.0	0.644	0.053
118	271	268	11.8	0.656	0.038
119	287	287	11.1	0.733	0.042
120	282	279	11.9	0.899	0.041
121	263	262	12.9	0.914	0.049
122	286	286	10.9	0.656	0.044
123	272	271	12.0	0.693	0.042
124	288	286	12.3	0.721	0.039
125	189	183	13.0	0.520	0.030
126	0	0	0.0	0.000	0.000
127	0	0	0.0	0.000	0.000
128	0	0	0.0	0.000	0.000
129	0	0	0.0	0.000	0.000
130	0	0	0.0	0.000	0.000

131	194	188	9.1	0.714	0.038
132	295	294	11.6	0.904	0.050
133	301	298	11.7	0.721	0.043
134	319	317	11.1	0.780	0.036
135	319	318	13.4	0.822	0.034
136	275	274	11.9	0.751	0.040
137	264	264	12.4	0.766	0.044
138	289	286	11.4	0.598	0.033
139	313	311	12.6	0.844	0.049
140	312	311	12.7	0.942	0.052
141	288	288	13.1	0.841	0.043
142	273	270	12.0	0.822	0.048
143	318	318	10.4	0.785	0.055
144	305	301	11.7	1.000	0.047
145	284	283	13.0	0.870	0.040
146	305	300	11.7	0.764	0.044
147	298	296	12.2	1.548	0.067
148	301	293	13.0	1.055	0.046
149	325	319	12.9	1.117	0.051
150	283	280	13.8	0.970	0.043
151	291	291	13.0	0.730	0.040
152	304	300	12.7	0.827	0.047
153	298	298	13.0	0.746	0.042
154	300	298	12.9	0.835	0.043
155	295	293	13.4	0.709	0.035
156	285	284	14.5	0.893	0.041
157	310	308	13.3	0.773	0.037
158	308	307	13.0	0.677	0.035
159	299	297	11.8	0.751	0.036
160	281	280	12.3	1.278	0.058
161	327	318	11.5	1.183	0.072
162	314	304	12.8	0.925	0.044
163	352	349	12.6	0.795	0.040
164	349	346	11.6	0.748	0.037
165	329	323	12.4	0.743	0.038
166	308	303	11.2	0.781	0.044
167	331	324	12.5	0.917	0.042
168	328	320	12.6	0.963	0.046
169	323	315	12.1	0.858	0.039
170	323	318	13.8	0.856	0.037
171	302	300	12.9	0.976	0.038
172	343	341	10.6	0.797	0.045
173	362	362	13.1	0.977	0.043
174	332	330	12.4	1.022	0.050
175	297	293	13.6	0.959	0.042
176	327	327	12.1	0.705	0.039
177	319	319	12.1	0.786	0.039
178	364	364	10.9	0.933	0.056
179	288	288	12.4	0.861	0.042
180	177	174	14.2	0.793	0.036
All films	51883	51428	12.0	0.943	0.047

Summary of observation redundancies by shells:

Shell Lower limit	Shell Upper limit	No. of reflections with given No. of observations: 0	1	2	3	4	5-6	7-8	9-12	13-19	>19	total
50.00	3.21	67	40	65	115	190	95	720	0	0	0	1225
3.21	2.55	10	21	39	72	123	87	850	0	0	0	1192
2.55	2.23	1	9	37	102	75	98	861	0	0	0	1182
2.23	2.02	3	6	29	123	46	106	858	0	0	0	1168
2.02	1.88	6	4	39	93	74	149	803	0	0	0	1162
1.88	1.77	55	28	81	87	132	248	519	0	0	0	1095
1.77	1.68	259	102	124	105	157	222	187	0	0	0	897
1.68	1.61	584	126	96	79	79	115	69	0	0	0	564
1.61	1.54	823	61	58	32	72	75	11	0	0	0	309
1.54	1.49	1021	58	43	32	5	0	0	0	0	0	138
All hkl		2829	455	611	840	953	1195	4878	0	0	0	8932

Average Redundancy Per Shell

Shell Lower limit	Shell Upper limit	
50.00	3.21	6.0
3.21	2.55	6.5
2.55	2.23	6.6
2.23	2.02	6.7
2.02	1.88	6.6
1.88	1.77	5.7
1.77	1.68	4.3
1.68	1.61	3.6
1.61	1.54	3.4
1.54	1.49	1.9
All hkl		5.8

Shell Summary of observation redundancies:

Lower limit	Upper limit	% of reflections with given No. of observations: 0	1	2	3	4	5-6	7-8	9-12	13-19	>19	total
50.00	3.21	5.2	3.1	5.0	8.9	14.7	7.4	55.7	0.0	0.0	0.0	94.8
3.21	2.55	0.8	1.7	3.2	6.0	10.2	7.2	70.7	0.0	0.0	0.0	99.2
2.55	2.23	0.1	0.8	3.1	8.6	6.3	8.3	72.8	0.0	0.0	0.0	99.9
2.23	2.02	0.3	0.5	2.5	10.5	3.9	9.1	73.3	0.0	0.0	0.0	99.7
2.02	1.88	0.5	0.3	3.3	8.0	6.3	12.8	68.8	0.0	0.0	0.0	99.5
1.88	1.77	4.8	2.4	7.0	7.6	11.5	21.6	45.1	0.0	0.0	0.0	95.2
1.77	1.68	22.4	8.8	10.7	9.1	13.6	19.2	16.2	0.0	0.0	0.0	77.6
1.68	1.61	50.9	11.0	8.4	6.9	6.9	10.0	6.0	0.0	0.0	0.0	49.1
1.61	1.54	72.7	5.4	5.1	2.8	6.4	6.6	1.0	0.0	0.0	0.0	27.3
1.54	1.49	88.1	5.0	3.7	2.8	0.4	0.0	0.0	0.0	0.0	0.0	11.9
All hkl		24.1	3.9	5.2	7.1	8.1	10.2	41.5	0.0	0.0	0.0	75.9

Shell I/Sigma in resolution shells:

Lower limit	Upper limit	No. of reflections with I / Sigma less than: 0	1	2	3	5	10	20	>20	total
50.00	3.21	2	8	21	27	36	55	85	1140	1225
3.21	2.55	10	28	39	47	59	96	204	988	1192
2.55	2.23	10	32	60	90	147	272	506	676	1182
2.23	2.02	25	77	137	190	286	476	765	403	1168
2.02	1.88	65	172	280	365	516	752	995	167	1162
1.88	1.77	71	241	408	513	670	861	1019	76	1095
1.77	1.68	104	327	470	552	643	746	833	64	897
1.68	1.61	85	253	329	376	414	481	529	35	564
1.61	1.54	49	134	199	228	258	283	305	4	309
1.54	1.49	28	82	117	128	137	138	138	0	138
All hkl		449	1354	2060	2516	3166	4160	5379	3553	8932

Shell I/Sigma in resolution shells:

Lower limit	Upper limit	% of of reflections with I / Sigma less than: 0	1	2	3	5	10	20	>20	total
50.00	3.21	0.2	0.6	1.6	2.1	2.8	4.3	6.6	88.2	94.8
3.21	2.55	0.8	2.3	3.2	3.9	4.9	8.0	17.0	82.2	99.2
2.55	2.23	0.8	2.7	5.1	7.6	12.4	23.0	42.8	57.1	99.9
2.23	2.02	2.1	6.6	11.7	16.2	24.4	40.6	65.3	34.4	99.7
2.02	1.88	5.6	14.7	24.0	31.2	44.2	64.4	85.2	14.3	99.5
1.88	1.77	6.2	21.0	35.5	44.6	58.3	74.9	88.6	6.6	95.2
1.77	1.68	9.0	28.3	40.7	47.8	55.6	64.5	72.1	5.5	77.6
1.68	1.61	7.4	22.0	28.7	32.8	36.1	41.9	46.1	3.0	49.1
1.61	1.54	4.3	11.8	17.6	20.1	22.8	25.0	26.9	0.4	27.3
1.54	1.49	2.4	7.1	10.1	11.0	11.8	11.9	11.9	0.0	11.9
All hkl		3.8	11.5	17.5	21.4	26.9	35.4	45.7	30.2	75.9

Summary of reflections intensities and R-factors by intensity bins
R linear = SUM (ABS(I - <I>)) / SUM (I)
R square = SUM ((I - <I>) ** 2) / SUM (I ** 2)
Chi**2 = SUM ((I - <I>) ** 2) / (Error ** 2 * N / (N-1)))
In all sums single measurements are excluded

Number	Average I	Average error	Average stat.	Norm. Chi**2	Linear R-fac	Square R-fac	nonunf
1496	-15.7	76.1	76.1	0.477	0.000	0.000	0.000
9913	49.0	76.0	76.0	0.468	0.770	0.954	0.000
5232	149.0	80.8	80.6	0.516	0.282	0.364	0.000
5940	292.9	86.9	86.5	0.565	0.163	0.206	0.000
5570	586.2	98.6	97.0	0.698	0.102	0.132	0.000
5412	1154.0	121.2	116.1	0.856	0.070	0.094	0.000
4628	2302.8	169.7	154.7	1.073	0.055	0.082	0.000
3851	4590.5	265.0	225.8	1.324	0.047	0.061	0.000
3042	9044.0	444.6	351.1	1.692	0.046	0.108	0.000
2124	18198.8	789.5	567.4	1.877	0.042	0.055	0.000
1682	36720.7	1459.1	951.7	2.300	0.042	0.057	0.000
1321	72361.4	2692.0	1579.7	2.363	0.040	0.053	0.000
846	142489.7	5016.2	2603.5	2.480	0.037	0.049	0.000
351	263423.2	8926.2	4089.5	2.295	0.034	0.045	0.000
20	476405.4	15639.6	6158.5	5.119	0.047	0.058	0.000

Summary of reflections intensities and R-factors by shells
R linear = SUM (ABS(I - <I>)) / SUM (I)
R square = SUM ((I - <I>) ** 2) / SUM (I ** 2)
Chi**2 = SUM ((I - <I>) ** 2) / (Error ** 2 * N / (N-1)))
In all sums single measurements are excluded

Shell limit	Lower Angstrom	Upper	Average I	Average error	Average stat.	Norm. Chi**2	Linear R-fac	Square R-fac
50.00	3.21	15844.0	281.4	176.9	2.165	0.041	0.053	
3.21	2.55	3781.5	79.2	57.9	1.230	0.043	0.056	
2.55	2.23	1089.9	34.0	30.1	0.891	0.058	0.063	
2.23	2.02	513.1	27.4	26.0	0.689	0.086	0.081	

2.02	1.88	262.2	25.0	24.5	0.594	0.144	0.113
1.88	1.77	172.5	27.3	26.9	0.529	0.200	0.148
1.77	1.68	181.7	33.3	32.8	0.525	0.171	0.107
1.68	1.61	163.9	36.8	36.3	0.511	0.150	0.097
1.61	1.54	89.6	31.5	31.4	0.474	0.272	0.185
1.54	1.49	36.0	41.5	41.5	0.445	0.558	0.491
All reflections		2976.4	71.2	53.2	0.959	0.047	0.053

Intensities of systematic absences

h	k	l	Intensity	Sigma	I/Sigma
0	0	3	4.7	1.5	3.2
0	0	5	1.7	2.5	0.7
0	0	7	4.2	2.2	1.9
0	0	9	5.6	2.9	1.9
0	0	11	8.1	3.4	2.4
0	0	13	7.0	4.1	1.7
0	0	15	25.2	6.2	4.1
0	0	17	22.8	10.5	2.2
0	0	19	30.6	21.8	1.4
0	0	23	-0.6	8.2	-0.1
0	0	25	2.6	7.3	0.4
0	0	27	13.4	7.5	1.8
0	0	29	-6.2	7.5	-0.8
0	3	0	4.7	5.1	0.9
0	5	0	0.4	6.5	0.1
0	7	0	78.7	16.2	4.9
0	9	0	-5.5	13.2	-0.4
0	11	0	54.3	29.5	1.8
0	13	0	-12.4	29.2	-0.4
0	15	0	5.4	43.3	0.1
0	17	0	22.0	31.0	0.7
0	19	0	28.2	57.7	0.5
0	21	0	-5.6	32.4	-0.2
0	23	0	-4.7	55.8	-0.1
1	0	0	10.3	3.4	3.1
3	0	0	8.9	10.1	0.9
5	0	0	5.6	17.5	0.3
7	0	0	245.5	86.8	2.8
9	0	0	22.8	39.7	0.6
11	0	0	6.7	45.9	0.1

13 0 0 -36.1 55.9 -0.6

Statistics for $^{TX}DDD^{2+}$ Structure (Crystal tm1605a)

number of film packs 180
total number of reflections used 115580
Goniostat axis directions:

phi	0.00000	0.00000	-1.00000
kappa	-1.00000	0.00000	0.00000
omega	0.00000	1.00000	0.00000

Goniostat angles:

o-detector,	o-beam,	o-zero,	o-kappa,	k-zero,	phi-kappa
0.000	0.000	-90.000	90.000	-90.000	90.000

Space group p212121 number 19 number of symmetry operators 4
x, y, z;
1/2-x, -y, z+1/2;
x+1/2, 1/2-y, -z;
-x, y+1/2, 1/2-z;
Unit cell volume 66552.2

New scale

1	0.9760	2	0.9808	3	0.9724	4	0.9837	5	0.9953
6	1.0095	7	1.0000	8	1.0104	9	1.0231	10	1.0376
11	1.0218	12	1.0244	13	1.0605	14	1.0365	15	1.0238
16	1.0365	17	1.0535	18	1.0456	19	1.0416	20	1.0604
21	1.0734	22	1.0650	23	1.0558	24	1.0824	25	1.0843
26	1.1053	27	1.0972	28	1.0912	29	1.1109	30	1.1255
31	1.1202	32	1.1221	33	1.1384	34	1.1461	35	1.1356
36	1.1533	37	1.1605	38	1.1774	39	1.1677	40	1.1726
41	1.2082	42	1.2129	43	1.2166	44	1.2344	45	1.2749
46	1.2641	47	1.2677	48	1.2603	49	1.2959	50	1.2960
51	1.3119	52	1.3028	53	1.3142	54	1.3332	55	1.3116
56	1.3066	57	1.3395	58	1.3159	59	1.3115	60	1.3110

61	1.3238	62	1.3191	63	1.2919	64	1.2903	65	1.2904
66	1.2989	67	1.2773	68	1.2786	69	1.2867	70	1.2676
71	1.2620	72	1.2438	73	1.2501	74	1.2301	75	1.2326
76	1.2226	77	1.2127	78	1.2029	79	1.1932	80	1.1835
81	1.1739	82	1.1588	83	1.1469	84	1.1494	85	1.1363
86	1.1222	87	1.0895	88	1.0746	89	1.0791	90	1.0688
91	1.0489	92	1.0206	93	1.0254	94	1.0023	95	0.9815
96	0.9715	97	0.9616	98	0.9518	99	0.9421	100	0.9325
101	0.9229	102	0.9135	103	0.9042	104	0.8950	105	0.8859
106	0.8768	107	0.8489	108	0.8352	109	0.8381	110	0.8228
111	0.8215	112	0.8218	113	0.8230	114	0.8338	115	0.8201
116	0.8126	117	0.8052	118	0.7978	119	0.7905	120	0.7832
121	0.7760	122	0.7798	123	0.7898	124	0.7805	125	0.7786
126	0.7869	127	0.7847	128	0.8085	129	0.7895	130	0.7779
131	0.7635	132	0.7683	133	0.7656	134	0.7778	135	0.7739
136	0.7662	137	0.7736	138	0.7823	139	0.7833	140	0.7737
141	0.7812	142	0.7736	143	0.7747	144	0.7765	145	0.7737
146	0.7619	147	0.7707	148	0.7759	149	0.7703	150	0.7700
151	0.7583	152	0.7689	153	0.7701	154	0.7684	155	0.7455
156	0.7665	157	0.7546	158	0.7529	159	0.7452	160	0.7470
161	0.7557	162	0.7564	163	0.7541	164	0.7574	165	0.7597
166	0.7557	167	0.7518	168	0.7455	169	0.7635	170	0.7561
171	0.7418	172	0.7575	173	0.7526	174	0.7638	175	0.7584
176	0.7607	177	0.7697	178	0.7796	179	0.7675	180	0.7803

Errors of the scale factors: film #, % error

1	0.0	2	0.0	3	0.0	4	0.0	5	0.0
6	0.0	8	0.0	9	0.0	10	0.0	11	0.0
12	0.0	13	0.0	14	0.0	15	0.0	16	0.0
17	0.0	18	0.0	19	0.0	20	0.0	21	0.0
22	0.0	23	0.0	24	0.0	25	0.0	26	0.0
27	0.0	28	0.0	29	0.0	30	0.0	31	0.0
32	0.0	33	0.0	34	0.0	35	0.0	36	0.0
37	0.0	38	0.0	39	0.0	40	0.0	41	0.0
42	0.0	43	0.0	44	0.0	45	0.0	46	0.0
47	0.0	48	0.0	49	0.0	50	0.0	51	0.0
52	0.0	53	0.0	54	0.0	55	0.0	56	0.0
57	0.0	58	0.0	59	0.0	60	0.0	61	0.0
62	0.0	63	0.0	64	0.0	65	0.0	66	0.0
67	0.0	68	0.0	69	0.0	70	0.0	71	0.0
72	0.0	73	0.0	74	0.0	75	0.0	76	0.0
77	0.0	78	0.0	79	0.0	80	0.0	81	0.0
82	0.0	83	0.0	84	0.0	85	0.0	86	0.0
87	0.0	88	0.0	89	0.0	90	0.0	91	0.0
92	0.0	93	0.0	94	0.0	95	0.0	96	0.0
97	0.0	98	0.0	99	0.0	100	0.0	101	0.0
102	0.0	103	0.0	104	0.0	105	0.0	106	0.0
107	0.0	108	0.0	109	0.0	110	0.0	111	0.0
112	0.0	113	0.0	114	0.0	115	0.0	116	0.0
117	0.0	118	0.0	119	0.0	120	0.0	121	0.0
122	0.0	123	0.0	124	0.0	125	0.0	126	0.0
127	0.0	128	0.0	129	0.0	130	0.0	131	0.0
132	0.0	133	0.0	134	0.0	135	0.0	136	0.0
137	0.0	138	0.0	139	0.0	140	0.0	141	0.0
142	0.0	143	0.0	144	0.0	145	0.0	146	0.0
147	0.0	148	0.0	149	0.0	150	0.0	151	0.0
152	0.0	153	0.0	154	0.0	155	0.0	156	0.0
157	0.0	158	0.0	159	0.0	160	0.0	161	0.0
162	0.0	163	0.0	164	0.0	165	0.0	166	0.0
167	0.0	168	0.0	169	0.0	170	0.0	171	0.0
172	0.0	173	0.0	174	0.0	175	0.0	176	0.0
177	0.0	178	0.0	179	0.0	180	0.0		

Summary of reflection intensities and R-factors by batch number

		All data		Linear
Batch	# obs # obs > 1	<I/sigma>	N. Chi**2	R-fac
1	353 353	13.9	1.607	0.060
2	418 417	13.4	1.202	0.054
3	407 407	14.8	1.206	0.047
4	396 396	13.3	1.360	0.059
5	343 341	12.8	1.037	0.040
6	444 442	15.2	1.100	0.048
7	454 453	15.8	1.193	0.044
8	418 416	14.7	1.632	0.060
9	429 427	14.3	1.131	0.044
10	375 373	16.7	1.401	0.040
11	393 392	15.5	1.302	0.049
12	451 451	14.5	1.096	0.041
13	438 436	14.7	1.216	0.054
14	422 422	16.0	1.230	0.048
15	355 353	14.7	1.268	0.044

16	402	402	16.9	1.520	0.051
17	472	470	14.7	1.510	0.048
18	423	422	16.1	1.509	0.050
19	399	398	15.6	1.232	0.048
20	391	388	16.8	1.259	0.042
21	477	477	14.9	0.931	0.038
22	397	396	14.3	1.417	0.057
23	384	384	16.0	1.316	0.050
24	418	416	15.4	1.342	0.049
25	436	436	14.7	0.986	0.042
26	415	414	15.0	0.962	0.041
27	430	430	15.2	1.155	0.041
28	379	378	15.0	1.006	0.037
29	425	425	14.8	1.394	0.044
30	418	417	14.0	0.958	0.037
31	400	400	16.5	1.464	0.049
32	407	407	16.1	1.100	0.035
33	432	428	15.6	1.234	0.048
34	400	395	13.8	1.172	0.049
35	400	400	16.5	1.306	0.053
36	410	408	14.6	0.984	0.046
37	405	403	15.6	1.042	0.039
38	409	406	14.1	1.135	0.053
39	425	423	15.8	0.927	0.038
40	390	388	16.1	1.161	0.043
41	404	403	15.0	1.118	0.039
42	442	439	14.7	1.110	0.045
43	407	406	16.0	1.047	0.039
44	413	413	16.0	0.912	0.033
45	437	436	15.2	1.028	0.037
46	406	401	16.4	1.043	0.037
47	416	414	16.8	0.979	0.038
48	460	459	16.3	1.132	0.043
49	391	388	16.1	1.040	0.040
50	407	402	14.9	0.760	0.034
51	392	389	15.4	0.959	0.038
52	418	415	15.0	0.944	0.043
53	454	452	16.9	0.965	0.034
54	403	400	15.0	0.699	0.033
55	396	395	16.3	1.021	0.038
56	406	403	15.1	1.248	0.046
57	425	423	16.1	0.784	0.034
58	440	436	17.4	0.807	0.032
59	410	407	16.8	0.953	0.036
60	391	387	14.9	0.944	0.039
61	430	428	14.5	0.881	0.039
62	404	398	14.4	0.794	0.038
63	421	419	15.7	0.940	0.044
64	426	424	15.3	0.938	0.038
65	424	421	14.4	0.858	0.042
66	407	404	16.0	0.849	0.035
67	425	422	16.1	0.887	0.039
68	412	408	16.0	0.877	0.035
69	384	379	14.9	0.823	0.038
70	437	428	15.1	0.818	0.035
71	399	396	15.9	0.894	0.040
72	420	417	14.3	0.849	0.037
73	426	425	17.0	0.943	0.039
74	415	408	14.8	1.021	0.042
75	322	320	15.9	0.911	0.035
76	0	0	0.0	0.000	0.000
77	0	0	0.0	0.000	0.000
78	0	0	0.0	0.000	0.000
79	0	0	0.0	0.000	0.000
80	0	0	0.0	0.000	0.000
81	315	312	13.9	0.757	0.033
82	416	413	14.8	1.107	0.044
83	413	407	15.3	0.987	0.039
84	426	425	15.5	0.976	0.045
85	433	431	15.8	1.691	0.038
86	419	415	15.5	0.954	0.034
87	389	383	14.3	0.849	0.036
88	405	404	15.0	0.959	0.037
89	410	405	14.4	0.791	0.031
90	435	433	15.0	0.921	0.040
91	414	413	15.2	0.812	0.030
92	401	398	14.5	0.732	0.037
93	377	375	15.2	0.829	0.038
94	446	445	15.0	0.870	0.037
95	325	325	13.8	0.742	0.033
96	0	0	0.0	0.000	0.000

97	0	0	0.0	0.000	0.000
98	0	0	0.0	0.000	0.000
99	0	0	0.0	0.000	0.000
100	0	0	0.0	0.000	0.000
101	0	0	0.0	0.000	0.000
102	0	0	0.0	0.000	0.000
103	0	0	0.0	0.000	0.000
104	0	0	0.0	0.000	0.000
105	0	0	0.0	0.000	0.000
106	291	291	14.1	0.701	0.035
107	444	443	13.6	1.103	0.043
108	402	402	14.3	1.367	0.054
109	387	387	14.9	1.073	0.041
110	431	431	13.9	1.231	0.050
111	374	374	13.2	1.016	0.047
112	413	413	12.9	0.749	0.037
113	384	383	12.9	0.834	0.041
114	404	403	14.7	0.909	0.043
115	303	302	13.3	0.862	0.045
116	0	0	0.0	0.000	0.000
117	0	0	0.0	0.000	0.000
118	0	0	0.0	0.000	0.000
119	0	0	0.0	0.000	0.000
120	0	0	0.0	0.000	0.000
121	324	323	12.5	1.285	0.049
122	410	410	15.2	1.734	0.059
123	415	415	14.3	1.068	0.043
124	443	443	13.3	0.994	0.042
125	391	390	13.3	1.487	0.052
126	401	400	14.0	1.060	0.043
127	402	402	13.7	1.322	0.054
128	406	406	13.7	3.021	0.091
129	429	428	12.9	1.252	0.051
130	389	388	14.3	1.543	0.057
131	388	387	14.8	1.510	0.059
132	402	401	14.6	1.419	0.053
133	478	478	12.5	1.259	0.052
134	396	396	14.7	1.768	0.069
135	356	356	12.8	1.669	0.053
136	394	393	12.8	1.918	0.076
137	429	429	13.7	1.684	0.061
138	456	455	13.6	1.579	0.062
139	370	369	14.8	1.706	0.065
140	359	359	14.5	1.505	0.061
141	399	398	13.0	1.802	0.069
142	422	421	14.8	1.531	0.058
143	434	434	15.1	1.409	0.055
144	442	441	15.2	1.429	0.050
145	374	373	13.7	1.774	0.063
146	397	397	13.0	1.734	0.071
147	381	381	13.8	1.321	0.057
148	405	402	14.8	1.711	0.068
149	411	411	13.2	2.345	0.074
150	397	397	13.6	1.346	0.058
151	466	465	14.5	1.511	0.061
152	391	388	15.9	1.346	0.052
153	413	413	13.1	1.579	0.063
154	393	393	13.1	1.442	0.057
155	388	387	12.9	1.853	0.063
156	400	398	13.7	1.732	0.069
157	421	421	13.4	1.654	0.071
158	413	413	13.9	2.068	0.071
159	421	421	13.9	1.295	0.050
160	417	416	14.6	2.236	0.069
161	418	418	14.6	1.436	0.057
162	413	411	14.7	1.282	0.057
163	419	419	14.9	1.942	0.073
164	414	412	14.5	2.017	0.068
165	414	413	14.4	1.219	0.044
166	422	418	14.2	1.812	0.061
167	405	403	13.0	1.985	0.061
168	428	424	15.0	2.699	0.079
169	426	425	13.9	1.885	0.068
170	418	414	13.8	2.311	0.074
171	405	404	14.3	1.640	0.060
172	413	412	13.7	1.692	0.078
173	401	400	12.5	1.263	0.057
174	422	419	13.6	1.721	0.069
175	388	387	13.4	1.407	0.065
176	398	397	13.1	1.622	0.064
177	384	384	12.4	1.405	0.059

178	440	439	13.7	1.280	0.055	
179	429	428	15.8	1.202	0.049	
180	343	343	15.0	1.096	0.044	
All films	65190	64936	14.7	1.267	0.050	

Shell Summary of observation redundancies by shells:

Lower limit	Upper limit	No. of reflections with given No. of observations 0	1	2	3	4	5-6	7-8	9-12	13-19	>19	total
50.00	3.21	108	42	89	106	189	300	428	0	0	0	1154
3.21	2.55	10	7	34	66	149	289	620	0	0	0	1165
2.55	2.23	0	0	44	44	126	325	621	0	0	0	1160
2.23	2.02	0	0	33	58	103	332	617	0	0	0	1143
2.02	1.88	0	0	34	50	96	359	596	0	0	0	1135
1.88	1.77	0	0	30	49	96	362	599	0	0	0	1136
1.77	1.68	0	0	29	48	91	361	590	0	0	0	1119
1.68	1.61	0	0	29	62	89	363	585	0	0	0	1128
1.61	1.54	1	7	42	74	127	386	490	0	0	0	1126
1.54	1.49	213	198	229	191	158	110	5	0	0	0	891
All hkl		332	254	593	748	1224	3187	5151	0	0	0	11157

Shell Average Redundancy Per Shell

Lower limit	Upper limit	
50.00	3.21	5.4
3.21	2.55	6.1
2.55	2.23	6.2
2.23	2.02	6.3
2.02	1.88	6.3
1.88	1.77	6.3
1.77	1.68	6.3
1.68	1.61	6.2
1.61	1.54	5.9
1.54	1.49	2.8
All hkl		5.8

Shell Summary of observation redundancies:

Lower limit	Upper limit	% of reflections with given No. of observations 0	1	2	3	4	5-6	7-8	9-12	13-19	>19	total
50.00	3.21	8.6	3.3	7.1	8.4	15.0	23.8	33.9	0.0	0.0	0.0	91.4
3.21	2.55	0.9	0.6	2.9	5.6	12.7	24.6	52.8	0.0	0.0	0.0	99.1
2.55	2.23	0.0	0.0	3.8	3.8	10.9	28.0	53.5	0.0	0.0	0.0	100.0
2.23	2.02	0.0	0.0	2.9	5.1	9.0	29.0	54.0	0.0	0.0	0.0	100.0
2.02	1.88	0.0	0.0	3.0	4.4	8.5	31.6	52.5	0.0	0.0	0.0	100.0
1.88	1.77	0.0	0.0	2.6	4.3	8.5	31.9	52.7	0.0	0.0	0.0	100.0
1.77	1.68	0.0	0.0	2.6	4.3	8.1	32.3	52.7	0.0	0.0	0.0	100.0
1.68	1.61	0.0	0.0	2.6	5.5	7.9	32.2	51.9	0.0	0.0	0.0	100.0
1.61	1.54	0.1	0.6	3.7	6.6	11.3	34.3	43.5	0.0	0.0	0.0	99.9
1.54	1.49	19.3	17.9	20.7	17.3	14.3	10.0	0.5	0.0	0.0	0.0	80.7
All hkl		2.9	2.2	5.2	6.5	10.7	27.7	44.8	0.0	0.0	0.0	97.1

Shell I/Sigma in resolution shells:

Lower limit	Upper limit	No. of reflections with I / Sigma less than 0	1	2	3	5	10	20	>20	total
50.00	3.21	3	9	11	16	21	28	52	1102	1154
3.21	2.55	3	11	17	23	28	55	106	1059	1165
2.55	2.23	7	20	24	29	50	95	184	976	1160
2.23	2.02	7	21	29	39	62	120	243	900	1143
2.02	1.88	6	24	45	71	107	194	415	720	1135
1.88	1.77	11	44	74	105	161	293	552	584	1136
1.77	1.68	19	65	107	142	223	394	697	422	1119
1.68	1.61	26	88	177	241	377	638	911	217	1128
1.61	1.54	56	182	303	418	581	852	1047	79	1126
1.54	1.49	54	197	377	527	672	831	884	7	891
All hkl		192	661	1164	1611	2282	3500	5091	6066	11157

Shell I/Sigma in resolution shells:

Lower limit	Upper limit	% of of reflections with I / Sigma less than 0	1	2	3	5	10	20	>20	total
50.00	3.21	0.2	0.7	0.9	1.3	1.7	2.2	4.1	87.3	91.4
3.21	2.55	0.3	0.9	1.4	2.0	2.4	4.7	9.0	90.1	99.1
2.55	2.23	0.6	1.7	2.1	2.5	4.3	8.2	15.9	84.1	100.0
2.23	2.02	0.6	1.8	2.5	3.4	5.4	10.5	21.3	78.7	100.0
2.02	1.88	0.5	2.1	4.0	6.3	9.4	17.1	36.6	63.4	100.0
1.88	1.77	1.0	3.9	6.5	9.2	14.2	25.8	48.6	51.4	100.0
1.77	1.68	1.7	5.8	9.6	12.7	19.9	35.2	62.3	37.7	100.0
1.68	1.61	2.3	7.8	15.7	21.4	33.4	56.6	80.8	19.2	100.0
1.61	1.54	5.0	16.1	26.9	37.1	51.6	75.6	92.9	7.0	99.9
1.54	1.49	4.9	17.8	34.1	47.7	60.9	75.3	80.1	0.6	80.7
All hkl		1.7	5.8	10.1	14.0	19.9	30.5	44.3	52.8	97.1

Summary of reflections intensities and R-factors by intensity bins
R linear = SUM (ABS(I - <I>)) / SUM (I)

R square = SUM ((I - <I>) ** 2) / SUM (I ** 2)
Chi**2 = SUM ((I - <I>) ** 2) / (Error ** 2 * N / (N-1)))
In all sums single measurements are excluded

Number	Average I	Average error	Average stat.	Norm. Chi**2	Linear R-fac	Square R-fac	nonunf
747	-11.6	71.0	71.0	0.500	0.000	0.000	0.000
7726	50.4	69.5	69.5	0.503	0.709	0.833	0.000
5510	159.8	73.9	73.8	0.697	0.305	0.000	0.000
7132	294.5	79.0	78.5	0.663	0.153	0.199	0.000
7877	587.1	89.7	87.9	0.925	0.103	0.134	0.000
7777	1163.9	113.8	108.2	1.118	0.073	0.095	0.000
7386	2318.1	163.5	147.8	1.454	0.061	0.093	0.000
6476	4570.6	258.5	218.7	1.738	0.053	0.069	0.000
4863	9048.2	441.9	346.8	1.909	0.047	0.061	0.000
3299	18181.6	793.1	572.1	2.114	0.045	0.059	0.000
2571	36216.7	1449.5	951.5	2.492	0.044	0.059	0.000
1882	72313.9	2714.0	1619.6	2.703	0.042	0.055	0.000
1288	142555.6	5038.3	2637.8	2.962	0.041	0.054	0.000
379	263344.6	8994.4	4152.3	3.855	0.043	0.057	0.000
22	479467.7	16396.3	6431.7	7.027	0.062	0.076	0.000
1	853281.1	31071.7	8690.9	24.747	0.165	0.142	0.000

Summary of reflections intensities and R-factors by shells
R linear = SUM (ABS(I - <I>)) / SUM (I)
R square = SUM ((I - <I>) ** 2) / SUM (I ** 2)
Chi**2 = SUM ((I - <I>) ** 2) / (Error ** 2 * N / (N-1)))
In all sums single measurements are excluded

Shell limit	Lower Angstrom	Upper	Average I	Average error	Average stat.	Norm. Chi**2	Linear R-fac	Square R-fac
50.00	3.21		30204.8	565.5	351.4	2.765	0.044	0.056
3.21	2.55		10244.6	189.2	128.8	1.903	0.046	0.072
2.55	2.23		3984.8	83.8	64.7	1.543	0.050	0.064
2.23	2.02		2754.9	66.0	54.1	1.335	0.053	0.066
2.02	1.88		1598.4	49.1	43.3	1.040	0.060	0.065
1.88	1.77		1119.2	42.3	38.7	1.029	0.081	0.529
1.77	1.68		856.7	39.5	36.7	0.884	0.086	0.085
1.68	1.61		610.4	36.6	34.6	0.739	0.110	0.085
1.61	1.54		286.8	34.8	34.2	0.697	0.187	0.172
1.54	1.49		202.3	58.9	58.7	0.639	0.238	0.226
All reflections			5359.7	118.9	85.7	1.289	0.050	0.063

Intensities of systematic absences

h	k	l	Intensity	Sigma	I/Sigma
0	0	3	20.1	5.2	3.8
0	0	5	12.3	2.6	4.7
0	0	7	9.2	2.8	3.3
0	0	9	7.4	4.5	1.6
0	0	11	3.6	5.2	0.7
0	0	13	2.4	7.7	0.3
0	0	15	29.4	10.1	2.9
0	0	17	5.1	14.3	0.4
0	0	19	-4.9	14.1	-0.3
0	0	21	3.3	16.5	0.2
0	0	23	8.1	17.4	0.5
0	0	25	2.9	13.6	0.2
0	0	27	25.7	15.9	1.6
0	0	29	9.8	18.3	0.5
0	0	31	15.7	18.3	0.9
0	0	33	-9.4	20.1	-0.5
0	0	35	24.9	21.1	1.2
0	0	37	0.9	19.7	0.0
0	0	39	-12.8	22.6	-0.6
0	0	41	3.3	27.9	0.1
0	0	43	0.8	28.5	0.0
0	3	0	13.0	6.0	2.2
0	5	0	27.0	12.7	2.1
0	7	0	8.2	12.4	0.7
0	9	0	4.0	22.8	0.2
0	11	0	-7.8	24.3	-0.3
0	13	0	-3.4	28.2	-0.1
0	15	0	-19.2	33.6	-0.6
0	17	0	-0.6	50.2	0.0
0	19	0	6.7	49.8	0.1
0	21	0	20.7	49.4	0.4
0	23	0	27.9	66.3	0.4
0	25	0	4.0	43.4	0.1
1	0	0	19.6	4.0	4.9
3	0	0	15.9	13.5	1.2
5	0	0	6.7	12.3	0.5
7	0	0	-3.3	20.1	-0.2

9	0	0	24.6	35.7	0.7
11	0	0	26.4	34.4	0.8
13	0	0	36.0	45.0	0.8
15	0	0	-31.4	43.3	-0.7

APPENDIX II

SAMPLE PERL PROGRAMS

Contact

```
#!/usr/bin/perl

##THIS PROGRAM DOES NOT GENERATE SYMMETRY.  TO GENERATE SYMMETRY USE “neighbors.inp” FROM
##CNS AND COMBINE THE OUTPUT SYMMTERY RELATED STRUCTURES WITH THE ORIGINAL MODEL USING
##”merge_structures.inp”.  THEN USE THIS PROGRAM

##set input and output files
$in = “PUT YOUR FILE NAME HERE";
$out = "PUT YOUR FILR NAME HERE";
##set distance
$lower = 0.0;
$upper = 25;
$verbose = 0;

open STDIN, "<$in" or die "Input file FAILED to open: $in";
open OUTFILE, ">$out" or die "Output file FAILED to open: $out"
print $out."\n";
print "Distance Lower Limit: $lower\n";
print "Distance Upper Limit: $upper\n\n";

$format = " %5s  %6d  %-7s  %-7s | %5s  %6d  %-7s  %-7s | %11.6f \n";
print "Parsing the first set of atoms from the file...\n";
$atomsA = 0;
while(defined($_ = <STDIN>)){
  chomp($_);
  if(/^(ATOM|HETATM) +\d+ +(P) +([A-Z]+) +(\d+) +([\d.\-]+) +([\d.\-]+) +([\d.\-]+).+ (M)
*$/){
    print $_."\n";
    push @AatmName, $2;
    push @AresName, $3;
    push @AresNum , $4;
    push @Ax      , $5;
    push @Ay      , $6;
    push @Az      , $7;
    push @AsegID  , $8;
    $atomsA++;
  }#if
}#while
print "$atomsA atoms selected for set A.\n";
close STDIN;
print "Preparing output file header...\n";
select OUTFILE;
print "An brute-force program written by Tinoush Moulaei on Friday, August 13th 2004.\n";
print "So that I may leave this place and live a normal life...\n";
print "Input file opened: $in\n";
print "Output file opened: $out\n";
print "Distance Lower Limit: $lower\n";
print "Distance Upper Limit: $upper\n\n";
print
"_____________________________________________________________________________\n";
  print "          set A atoms            |          set B atoms            |  Distance
\n";
  print " SegID  ResNum  ResName  AtmName | SegID  ResNum  ResName  AtmName | (angstroms)
\n";
  print
"_____________________________________________________________________________\n";
```

```
select STDOUT;
print "Parsing the second set of atoms from the file...\n";
print "Also calculating the distances to avoid a huge second array...\n";
if($verbose){
  print
"_____________________________________________________________________________________\n";
  print "          set A atoms            |          set B atoms            |  Distance
\n";
  print " SegID  ResNum  ResName  AtmName | SegID  ResNum  ResName  AtmName | (angstroms)
\n";
  print
"_____________________________________________________________________________________\n";
}#if
$atomB = 0;
while(defined($id = shift(@AsegID))){
  $aName = shift(@AatmName);
  $rName = shift(@AresName);
  $rNum  = shift(@AresNum);
  $x     = shift(@Ax);
  $y     = shift(@Ay);
  $z     = shift(@Az);
  open STDIN, "<$in";
  while(defined($_ = <STDIN>)){
    chomp($_);
    if(/^(ATOM|HETATM) +\d+ +(C5A) +([\dA-Z]+) +(\d+) +([\d.\-]+) +([\d.\-]+) +([\d.\-
]+).+ ([\dA-Z]+) *$/){
      $dist = sqrt(($x - $5)**2 + ($y - $6)**2 + ($z - $7)**2);
      if($dist <= $lower){if($verbose){print "-";}}#ifif
      elsif($dist > $upper){if($verbose){print "+";}}#elsifif
      else{printf OUTFILE $format, $id, $rNum, $rName, $aName, $8, $4, $3, $2,
$dist;}#else
      if($verbose){
        print "O";
        printf $format, $id, $rNum, $rName, $aName, $8, $4, $3, $2, $dist;
      }#if
      $atomsB++;
    }#if
  }#while
  close STDIN;
}#while
print
"_____________________________________________________________________________________\n";
$atomsB = $atomsB/$atomsA;
print "$atomsB atoms selected for set B.\n";
print "I hope that was as fun for you as it was for me.\n";
print "Remember to question your government if you want your grandchildren to have a good
life!\n";
select OUTFILE;
print "$atomsA atoms selected for set A.\n";
print "$atomsB atoms selected for set B.\n";
print "I hope that was as fun for you as it was for me.\n";
print "Remember to question your government if you want your grandchildren to have a good
life!\n";
close STDIN;
close OUTFILE;
```

Sort

```
#!/usr/bin/perl

##set input and output files
$in = "/home/monk/loren/NA/NaM1.pdb";
$out = "NaM1sort.pdb";
##set distance
$water = 27;
$sort = 1;
```

```
open STDIN, "<$in" or die "Input file FAILED to open: $in";
open OUTFILE, ">$out" or die "Output file FAILED to open: $out";
print "Numbering waters starting with $water.\n";
print "Waters will ";
if($sort){print "BE sorted.\n";}else{print "NOT BE sorted.\n";}

while(defined($_ = <STDIN>)){
  chomp($_);
  if(/^(ATOM +\d+ +O +HOH)( +\d+)( +[\d.\-]+ +[\d.\-]+ +[\d.\-]+
+\d\.\d\d)( +[\d.]+).*/){
    push @begin, $1;
    push @middl, $3;
    push @therm, $4;
  }#if
  elsif(!/^END/){
    print OUTFILE "$_\n";
  }#elsif
}#while
$tot = $#begin + 1;
if($sort){
  for($ol = 0; $ol < $tot; $ol++){
    for($il = $ol; $il < $tot; $il++){
      if($therm[$il] < $therm[$ol]){
        $temp = $middl[$il];
        $middl[$il] = $middl[$ol];
        $middl[$ol] = $temp;
        $temp = $therm[$il];
        $therm[$il] = $therm[$ol];
        $therm[$ol] = $temp;
      }#if
    }#inner for
  }#outer for
}#if
for($i = 0; $i < $tot; $i++){
  printf OUTFILE "%-20s%6d%34s%6.2f\n", $begin[$i], $water++,
$middl[$i], $therm[$i];
}#for
print OUTFILE "END\n";
close STDIN;
close OUTFILE;
```

BIBLIOGRAPHY

(1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **D50,** 760-763.

ALEXEEV, D. G., LIPANOV, A. A. & SKURATOVSKII, I. Y. (1987). Poly(dA)-poly(dT) is a B-type double helix with a distinctively narrow minor groove. *Nature* **325,** 821-823.

ALLEMANN, R. K. & EGLI, M. (1997). DNA recognition and bending. *Chem. Biol.* **4,** 643-650.

ANDERSON, C. F. & RECORD, M. T., JR. (1995). Salt-nucleic acid interactions. *Annu. Rev. Phys. Chem.* **46,** 657-700.

BADGER, J., KAPULSKY, A., GURSKY, O., BHYRAVBHATLA, B. & CASPAR, D. L. (1994a). Structure and selectivity of a monovalent cation binding site in cubic insulin crystals. *Biophys. J.* **66,** 286-92.

BADGER, J., LI, Y. & CASPAR, D. L. (1994b). Thallium Counterion Distribution in Cubic Insulin Crystals Determined from Anomalous X-ray Diffraction Data. *Proc. Natl. Acad. Sci. U.S.A.* **91,** 1224-1228.

BANCROFT, D., WILLIAMS, L. D., RICH, A. & EGLI, M. (1994). The Low Temperature Crystal Structure of the Pure-Spermine Forms of Z-DNA Reveals Binding of a Spermine Molecule in the Minor Groove. *Biochemistry* **33,** 1073-1086.

BARTENEV, V. N., GOLOVAMOV EU, I., KAPITONOVA, K. A., MOKULSKII, M. A., VOLKOVA, L. I. & SKURATOVSKII, I. Y. (1983). Structure of the B DNA cationic shell as revealed by an X-ray diffraction study of CsDNA. Sequence-specific cationic stabilization of B form DNA. *J. Mol. Biol.* **169,** 217-234.

BASU, S., RAMBO, R. P., STRAUSS-SOUKUP, J., CATE, J. H., FERRE-D'AMARE, A. R., STROBEL, S. A. & DOUDNA, J. A. (1998). A Specific Monovalent Metal Ion Integral to the AA Platform of the RNA Tetraloop Receptor. *Nat. Struct. Biol.* **5,** 986-992.

BASU, S., SZEWCZAK, A. A., COCCO, M. & STROBEL, S. A. (2000). Direct Detection of Monovalent Metal Ion Binding to a DNA G-quartet by 205Tl NMR. *J. Am. Chem. Soc.* **122,** 3240-3241.

BENNETT, M., KRAH, A., WIEN, F., GARMAN, E., MCKENNA, R., SANDERSON, M. & NEIDLE, S. (2000). A DNA-porphyrin minor-groove complex at atomic

resolution: the structural consequences of porphyrin ruffling. *Proc Natl Acad Sci U S A* **97,** 9476-81.

BERMAN, H. M., ZARDECKI, C. & WESTBROOK, J. (1998). The Nucleic Acid Database: A resource for nucleic acid science. *Acta Crystallogr D Biol Crystallogr* **54,** 1095-104.

BROWN, I. D. (1981). The Bond Valence Method: An Emperical Approach to Chemical Structure and Bonding. In *Structure and Bonding in Crystals*, vol. II, pp. 1-31.

BROWN, I. D. (1988). What Factors Determine Cation Coordination Numbers. *Acta Crystallogr.* **B44,** 545-553.

BROWN, I. D. (1992). Chemical and steric constraints in inorganic solids. *Acta Cryst.* **B48,** 553-572.

BROWN, I. D. (1996). VALENCE: A program for calculating bond valences. *Journal of Applied Crystallography* **29,** 479-480.

BRUKNER, I., SUSIC, S., DLAKIC, M., SAVIC, A. & PONGOR, S. (1994). Physiological concentration of magnesium ions induces a strong macroscopic curvature in GGGCCC-containing DNA. *J. Mol. Biol.* **236,** 26-32.

BRUNGER, A. T., ADAMS, P. D., CLORE, G. M., DELANO, W. L., GROS, P., GROSSE-KUNSTLEVE, R. W., JIANG, J. S., KUSZEWSKI, J., NILGES, M., PANNU, N. S., READ, R. J., RICE, L. M., SIMONSON, T. & WARREN, G. L. (1998). Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **54,** 905-921.

BUCKIN, V. A., KANKIYA, B. I., RENTZEPERIS, D. & MARKY, L. A. (1994). Mg2+ Recognizes the Sequence of DNA Through Its Hydration Shell. *J. Am. Chem. Soc.* **116,** 9423-9429.

BURKHOFF, A. M. & TULLIUS, T. D. (1987). The unusual conformation adopted by the adenine tracts in kinetoplast DNA. *Cell* **48,** 935-943.

CESARE MARINCOLA, F., DENISOV, V. P. & HALLE, B. (2004). Competitive Na(+) and Rb(+) binding in the minor groove of DNA. *J Am Chem Soc* **126,** 6739-50.

CHIU, T. K. & DICKERSON, R. E. (2000). 1 Å crystal structures of B-DNA reveal sequence-specific binding and groove-specific bending of DNA by magnesium and calcium. *J. Mol. Biol.* **301,** 915-45.

CHIU, T. K., KACZOR-GRZESKOWIAK, M. & DICKERSON, R. E. (1999). Absence of Minor Groove Monovalent Cations in the Crosslinked Dodecamer CGCGAATTCGCG. *J. Mol. Biol.* **292,** 589-608.

CLOWNEY, L., JAIN, S. C., SRINIVASAN, A. R., WESTBROOK, J., OLSON, W. K. & BERMAN, H. M. (1996). Geometric Parameters in Nucleic Acids: Nitrogenous Bases. *J. Am. Chem. Soc.* **118,** 509-518.

DANDE, P., LIANG, G. N., CHEN, F. X., ROBERTS, C., NELSON, M. G., HASHIMOTO, H., SWITZER, C. & GOLD, B. (1997). Regioselective effect of zwitterionic DNA substitutions on DNA alkylation: Evidence for a strong side chain orientational preference. *Biochemistry* **36,** 6024-6032.

DAUTER, Z. & ADAMIAK, D. A. (2001). Anomalous signal of phosphorus used for phasing DNA oligomer: importance of data redundancy. *Acta Crystallogr. D., Biol. Crystallogr.* **57,** 990-995.

DENG, J., XIONG, Y. & SUNDARALINGAM, M. (2001). X-ray analysis of an RNA tetraplex (UGGGGU)(4) with divalent Sr(2+) ions at subatomic resolution (0.61 A). *Proc Natl Acad Sci U S A* **98,** 13665-70.

DENISOV, V. P. & HALLE, B. (2000). Sequence-specific binding of counterions to B-DNA. *Proc. Natl. Acad. Sci. U.S.A.* **97,** 629-33.

DICKERSON, R. E. R.

DIEKMANN, S. & WANG, J. C. (1985). On the sequence determinants and flexibility of the kinetoplast DNA fragment with abnormal gel electrophoretic mobilities. *J. Mol. Biol.* **186,** 1-11.

DLAKIC, M. & HARRINGTON, R. E. (1995). Bending and torsional flexibility of G/C-rich sequences as determined by cyclization assays. *J. Biol. Chem.* **270,** 29945-52.

DREW, H. R. & DICKERSON, R. E. (1981). Structure of a B-DNA dodecamer. III. Geometry of Hydration. *J. Mol. Biol.* **151,** 535-556.

DREW, H. R., WING, R. M., TAKANO, T., BROKA, C., ITAKURA, K. & DICKERSON, R. E. (1981). Structure of a B-DNA Dodecamer. Conformation and Dynamics. *Proc. Natl. Acad. Sci. U.S.A.* **78,** 2179-2183.

EGLI, M., TERESHKO, V., TEPLOVA, M., MINASOV, G., JOACHIMIAK, A., SANISHVILI, R., WEEKS, C. M., MILLER, R., MAIER, M. A., AN, H., DAN COOK, P. & MANOHARAN, M. (1998). X-ray crystallographic analysis of the hydration of A- and B-form DNA at atomic resolution. *Biopolymers* **48,** 234-52.

EGLI, M., WILLIAMS, L. D., GAO, Q. & RICH, A. (1991). Structure of the Pure-Spermine Form of Z-DNA (Magnesium Free) at 1Å Resolution. *Biochemistry* **30,** 11388-11402.

FRANKLIN, R. E. & GOSLING, R. G. (1953). Molecular Configuration in Sodium Thymonucleate. *Nature* **171,** 740-741.

FRATINI, A. V., KOPKA, M. L., DREW, H. R. & DICKERSON, R. E. (1982). Reversible bending and helix geometry in a B-DNA dodecamer: CGCGAATTBrCGCG. *J. Biol. Chem.* **257,** 14686-14707.

FREDERICK, C. A., WILLIAMS, L. D., UGHETTO, G., VAN DER MAREL, G. A., VAN BOOM, J. H., RICH, A. & WANG, A. H.-J. (1990). Structural Comparison of Anti-Cancer Drug-DNA Complexes: Adriamycin and Daunomycin. *Biochemistry* **29,** 2538-2549.

GELBIN, A., SCHNEIDER, B., CLOWNEY, L., HSIEH, S.-H., OLSON, W. K. & BERMAN, H. M. (1996). Geometric Parameters in Nucleic Acids: Sugar and Phosphate Constitutents. *J. Am. Chem. Soc.* **118,** 519-529.

GESSNER, R. V., FREDERICK, C. A., QUIGLEY, G. J., RICH, A. & WANG, A. H.-J. (1989). The Molecular Stucture of the Left-handed Z-DNA Double Helix at 1.0 A Atomic Resolution. *J. Biol. Chem.* **264,** 7921-7935.

GILL, H. S. & EISENBERG, D. (2001). The Crystal Structure of Phosphinothricin in the Active Site of Glutamine Synthetase Illuminates the Mechanism of Enzymatic Inhibition. *Biochemistry* **40,** 1903-1912.

GOLD, B. (2002). Effect of cationic charge localization on DNA structure. *Biopolymers* **65,** 173-9.

GOODSELL, D. S., KOPKA, M. L., CASCIO, D. & DICKERSON, R. E. (1993). Crystal structure of CATGGCCATG and its implications for A-tract bending models. *Proc. Natl. Acad. Sci. U.S.A.* **90,** 2930-2934.

GRZESKOWIAK, K., YANAGI, K., PRIVE, G. G. & DICKERSON, R. E. (1991). The structure of B-helical C-G-A-T-C-G-A-T-C-G and comparison with C-C- A-A-C-G-T-T-G-G. The effect of base pair reversals. *J. Biol. Chem.* **266,** 8861-8883.

GURSKY, O., BADGER, J., LI, Y. & CASPAR, D. L. (1992a). Conformational changes in cubic insulin crystals in the pH range 7-11. *Biophys. J.* **63,** 1210-20.

GURSKY, O., LI, Y., BADGER, J. & CASPAR, D. L. (1992b). Monovalent cation binding to cubic insulin crystals. *Biophys. J.* **61,** 604-11.

HAGERMAN, K. R. & HAGERMAN, P. J. (1996). Helix rigidity of DNA: the meroduplex as an experimental paradigm. *J. Mol. Biol.* **260,** 207-23.

HAGERMAN, P. J. (1984). Evidence for the existence of stable curvature of DNA in solution. *Proc. Natl. Acad. Sci. U.S.A.* **81,** 4632-4636.

HAGERMAN, P. J. (1986). Sequence-directed curvature of DNA. *Nature* **321,** 449-450.

HAGERMAN, P. J. (1988). Flexibility of DNA. *Annu Rev Biophys Biophys Chem* **17,** 265-86.

HAGERMAN, P. J. (1990). Sequence-directed curvature of DNA. *Annu. Rev. Biochem.* **59,** 755-781.

HAMELBERG, D., MCFAIL-ISOM, L., WILLIAMS, L. D. & WILSON, W. D. (2000). The Flexible Structure of DNA: Ion Dependence of Minor-Groove Structure and Dynamics. *J. Am. Chem. Soc.* **122,** 10513-10520.

HAMELBERG, D., WILLIAMS, L. D. & WILSON, W. D. (2002). The Effect of Neutralized Phosphate Backbone on the Minor Groove Structure of B-DNA. *Nucleic Acids Res.* **30,** 3615-3623.

HARAN, T. E., KAHN, J. D. & CROTHERS, D. M. (1994). Sequence elements responsible for DNA curvature. *J. Mol. Biol.* **244,** 135-143.

HERSHKOVITZ, E., TANNENBAUM, E., HOWERTON, S. B., SHETH, A., TANNENBAUM, A. & WILLIAMS, L. D. (2003). Automated Identification of RNA Conformational Motifs: Theory and Application to the HM LSU 23S rRNA. *Nucleic Acids Res* **31,** 6249-6257.

HEYSTEK, L. E., ZHOU, H. Q., DANDE, P. & GOLD, B. (1998). Control over the localization of positive charge in DNA: The effect on duplex DNA and RNA stability. *J. Am. Chem. Soc.* **120,** 12165-12166.

HOWERTON, S. B. (2002). Structural Characterization of B-DNA and its Interactions with Cations and Intercalating Ligands. Ph.D. thesis, Georgia Institute of Technology.

HOWERTON, S. B., NAGPAL, A. & WILLIAMS, L. D. (2003). Surprising Roles for Electrostatic Interactions in DNA-Ligand Complexes. *Biopolymers* **69,** 87-99.

HOWERTON, S. B., SINES, C. C., VANDERVEER, D. & WILLIAMS, L. D. (2001). Locating Monovalent Cations in the Grooves of B-DNA. *Biochemistry* **40,** 10023-10031.

HUD, N. V. & FEIGON, J. (1997). Localization of divalent metal ions in the minor groove of DNA A-tracts. *J. Am. Chem. Soc.* **119,** 5756-5757.

HUD, N. V. & FEIGON, J. (2002). Characterization of divalent cation localization in the minor groove of the A(n)T(n) and T(n)A(n) DNA sequence elements by (1)H NMR spectroscopy and manganese(II). *Biochemistry* **41,** 9900-10.

HUD, N. V. & PLAVEC, J. (2003). A unified model for the origin of DNA sequence-directed curvature. *Biopolymers* **69,** 144-58.

HUD, N. V. & POLAK, M. (2001). DNA-cation interactions: The major and minor grooves are flexible ionophores. *Curr. Opin. Struct. Biol.* **11,** 293-301.

HUD, N. V., SCHULTZE, P. & FEIGON, J. (1998). Ammonium ion as an NMR probe for monovalent cation coordination sites of DNA quadruplexes. *J. Am. Chem. Soc.* **120,** 6403-6404.

HUD, N. V., SKLENAR, V. & FEIGON, J. (1999). Localization of ammonium ions in the minor groove of DNA duplexes in solution and the origin of DNA A-tract bending. *J. Mol. Biol.* **286,** 651-60.

IKEDA, K., NAGANO, K. & KAWAKAMI, K. (1993). Possible implications of Sp1-induced bending of DNA on synergistic activation of transcription. *Gene* **136,** 341-3.

JANIAK, C. (1997). (Organo)thallium (I) and (II) chemistry: syntheses, structures, properties and applications of subvalent thallium complexes with alkyl, cyclopentadienyl, arene or hydrotris(pyrazolyl)borate ligands. *Coordination Chemistry Reviews* **163,** 107-216.

JEFFREY, G. A. (1997). *An Introduction to Hydrogen Bonding*. Oxford University Press, New York.

JUO, Z. S., CHIU, T. K., LEIBERMAN, P. M., BAIKALOV, I., BERK, A. J. & DICKERSON, R. E. (1996). How proteins recognize the TATA box. *J. Mol. Biol.* **261,** 239-254.

KAHN, J. D. & CROTHERS, D. M. (1998). Measurement of the DNA bend angle induced by the catabolite activator protein using Monte Carlo simulation of cyclization kinetics. *Journal of Molecular Biology* **276,** 287-309.

KERPPOLA, T. K. & CURRAN, T. (1991). DNA bending by Fos and Jun: the flexible hinge model. *Science* **254,** 1210-4.

KIELKOPF, C. L., DING, S., KUHN, P. & REES, D. C. (2000). Conformational Flexibility of B-DNA at 0.74 A Resolution: d(CCAGTACTGG)(2). *J. Mol. Biol.* **296,** 787-801.

KOO, H.-S., DRAK, J., RICE, J. A. & CROTHERS, D. M. (1990). Determination of the extent of DNA bending by an adenine-thymine tract. *Biochemistry* **29,** 4227-4234.

KOO, H.-S., WU, H.-M. & CROTHERS, D. M. (1986). DNA bending at adenine-thymine tracts. *Nature* **320,** 501-506.

KOPKA, M. L., FRATINI, A. V., DREW, H. R. & DICKERSON, R. E. (1983). Ordered water structure around a B-DNA dodecamer. A quantitative study. *J. Mol. Biol.* **163,** 129-146.

KORNBERG, R. D. & KLUG, A. (1981). The nucleosome. *Sci Am* **244,** 52-64.

KOUDELKA, G. B., HARBURY, P., HARRISON, S. C. & PTASHNE, M. (1988). DNA Twisting and the Affinity of Bacteriophage 434 Operator for Bacteriophage 434 Repressor. *Proc. Natl. Acad. Sci. U.S.A.* **85,** 4633-4637.

KOUDELKA, G. B., HARRISON, S. C. & PTASHNE, M. (1987). Effect of Non-Contacted Bases on the Affinity of 434 Operator for 434 Repressor and Cro. *Nature* **326,** 886-888.

KRISTIANSSON, O. (2002). Structures of complexes of thallium(I) and functionalized benzoate ligands with pronounced stereoactivity of the lone pair of electrons and metal-phenyl pi-bonding. *European Journal of Inorganic Chemistry***,** 2355-2361.

LESLIE, A. G., ARNOTT, S., CHANDRASEKARAN, R. & RATLIFF, R. L. (1980). Polymorphism of DNA double helices. *J Mol Biol* **143,** 49-72.

LESLIE, A. G. W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography* **26**.

LI, Z., HUANG, L., DANDE, P., GOLD, B. & STONE, M. P. (2002). Structure of a tethered cationic 3-aminopropyl chain incorporated into an oligodeoxynucleotide: evidence for 3'-orientation in the major groove accompanied by DNA bending. *J Am Chem Soc* **124,** 8553-60.

LIANG, G. N., ENCELL, L., NELSON, M. G., SWITZER, C., SHUKER, D. E. G. & GOLD, B. (1995). Role of Electrostatics in the Sequence-Selective Reaction of Charged Alkylating-Agents With DNA. *J. Am. Chem. Soc.* **117,** 10135-10136.

LIPSCOMB, L. A., PEEK, M. E., ZHOU, F. X., BERTRAND, J. A., VANDERVEER, D. & WILLIAMS, L. D. (1994). Water Ring Structure at DNA Interfaces: Hydration and Dynamics of DNA-Anthracycline Complexes. *Biochemistry* **33,** 3649-3659.

LORIA, J. P. & NOWAK, T. (1998). Conformational Changes in Yeast Pyruvate Kinase Studied by 205Tl+ NMR. *Biochemistry* **37,** 6967-6974.

LU, X. J., SHAKKED, Z. & OLSON, W. K. (2000). A-form conformational motifs in ligand-bound DNA structures. *J Mol Biol* **300,** 819-40.

LUGER, K., MADER, A. W., RICHMOND, R. K., SARGENT, D. F. & RICHMOND, T. J. (1997). Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251-260.

MANNING, G. S. (1978). The molecular theory of polyelectrolyte solutions with applications to the electrostatic properties of polynucleotides. *Q. Rev. Biophys.* **11,** 179-246.

MANNING, G. S., EBRALIDSE, K. K., MIRZABEKOV, A. D. & RICH, A. (1989). An estimate of the extent of folding of nucleosomal DNA by laterally asymmetric neutralization of phosphate groups. *J. Biomol. Struct. Dyn.* **6,** 877-889.

MARINI, J. C., LEVENE, S. D., CROTHERS, D. M. & ENGLUND, P. T. (1982). Bent helical structure in kinetoplast DNA. *Proc. Natl. Acad. Sci., U.S.A.* **79,** 7664-7668.

MASQUIDA, B., SAUTER, C. & WESTHOF, E. (1999). A sulfate pocket formed by three GoU pairs in the 0.97 A resolution X-ray structure of a nonameric RNA. *Rna* **5,** 1384-95.

MCCONNELL, K. J. & BEVERIDGE, D. L. (2000). DNA structure: What's in charge? *J. Mol. Biol.* **304,** 803-20.

MCFAIL-ISOM, L., SINES, C. & WILLIAMS, L. D. (1999). DNA Structure: Cations in Charge? *Current Op. Struct. Biol.* **9,** 298-304.

MINASOV, G., TERESHKO, V. & EGLI, M. (1999). Atomic-resolution crystal structures of B-DNA reveal specific influences of divalent metal ions on conformation and packing. *J. Mol. Biol.* **291,** 83-99.

MIRZABEKOV, A. D. & RICH, A. (1979). Asymmetric lateral distribution of unshielded phosphate groups in nucleosomal DNA and its role in DNA bending. *Proc. Natl. Acad. Sci. U.S.A.* **76,** 1118-1121.

NAYAL, M. & DICERA, E. (1996). Valence screening of water in protein crystals reveals potential Na+ binding sites. *J. Mol. Biol.* **256,** 228-234.

NELSON, H. C. M., FINCH, J. T., LUISI, B. F. & KLUG, A. (1987). The Structure of an Oligo(dA)•Oligo(dT) Tract and its Biological Implications. *Nature* **330,** 221-226.

NG, H. L. & DICKERSON, R. E. (2002). Mediation of the A/B-DNA helix transition by G-tracts in the crystal structure of duplex CATGGGCCCATG. *Nucleic Acids Res* **30,** 4061-7.

NIKOLOV, D. B., CHEN, H., HALAY, E. D., HOFFMAN, A., ROEDER, R. G. & BURLEY, S. K. (1996). Crystal structure of a human TATA box-binding protein/TATA element complex. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 4862-4867.

NIKOLOV, D. B., CHEN, H., HALAY, E. D., USHEVA, A. A., HISATAKE, K., LEE, D. K., ROEDER, R. G. & BURLEY, S. K. (1995). Crystal structure of a TFIIB-TBP-TATA-element ternary complex. *Nature* **377,** 119-128.

OHISHI, H., KUNISAWA, S., VAN DER MAREL, G., VAN BOOM, J. H., RICH, A., WANG, A. H.-J., TOMITA, K. & HAKOSHIMA, T. (1991). Interaction Between the Left-Handed Z-DNA and Polyamine: The Crystal Structure of the $d(CG)_3$ and N-(2-aminoethyl)-1,4-diamino-butane Complex. *FEBS Let.* **284,** 238-244.

OHISHI, H., NAKANISHI, I., INUBUSHI, K., VAN DER MAREL, G., VAN BOOM, J. H., RICH, A., WANG, A. H., HAKOSHIMA, T. & TOMITA, K. (1996a). Interaction between the left-handed Z-DNA and polyamine-2. The crystal structure of the d(CG)3 and spermidine complex. *FEBS Lett* **391,** 153-6.

OHISHI, H., TERASOMA, N., NAKANISHI, I., VAN DER MAREL, G., VAN BOOM, J. H., RICH, A., WANG, A. H., HAKOSHIMA, T. & TOMITA, K. (1996b). Interaction between left-handed Z-DNA and polyamine - 3. The crystal structure of the d(CG)3 and thermospermine complex. *FEBS Lett* **398,** 291-296.

OLSON, W. K., GORIN, A. A., LU, X. J., HOCK, L. M. & ZHURKIN, V. B. (1998). DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. *Proc. Natl. Acad. Sci. U.S.A.* **95,** 11163-11168.

OTWINOWSKI, Z. (1993). Data Collection and Processing. In *Data Collection and Processing*, pp. 56-62. Science and Engineering Research Council, Warrington, United Kingdom.

OTWINOWSKI, Z. & MINOR, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymol., Macromolecular Crystallography*, vol. 276, Part A (ed. J. Carter, C.W. and R. M. Sweet), pp. 307-326. Academic Press, New York.

PAOLELLA, D. N., LIU, Y., FABIAN, M. A. & SCHEPARTZ, A. (1997). Electrostatic mechanism for DNA bending by bZIP proteins. *Biochemistry* **36,** 10033-10038.

PARKINSON, G., VOJTECHOVSKY, J., CLOWNEY, L., BRUNGER, A. T. & BERMAN, H. M. (1996). New parameters for the refinement of nucleic acid-containing structures. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **52,** 57-64.

PASSNER, J. M. & STEITZ, T. A. (1997). The structure of a CAP-DNA complex having two cAMP molecules bound to each monomer. *Proc. Natl. Acad. Sci. U.S.A.* **94,** 2843-7.

PEDERSEN, P. A., NIELSEN, J. M., RASMUSSEN, J. H. & JORGENSEN, P. L. (1998). Contribution to Tl+, K+, and Na+ binding of Asn776, Ser775, Thr774, Thr772, and Tyr771 in cytoplasmic part of fifth transmembrane segment in alpha-subunit of renal Na,K-ATPase. *Biochemistry* **37,** 17818-27.

PELLEGRINI, L., TAN, S. & RICHMOND, T. J. (1995). Structure of serum response factor core bound to DNA. *Nature* **376,** 490-498.

PHILLIPS, K., DAUTER, Z., MURCHIE, A. I., LILLEY, D. M. & LUISI, B. (1997). The crystal structure of a parallel-stranded guanine tetraplex at 0.95 A resolution. *J Mol Biol* **273,** 171-82.

POWELL, H. R. (1999). The Rossmann Fourier autoindexing algorithm in MOSFLM. *Acta Crystallogr. Sect. D-Biol. Crystallogr.* **55,** 1690-5.

RECORD, M. T., JR., ANDERSON, C. F. & LOHMAN, T. M. (1978). Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity. *Q Rev Biophys* **11,** 103-78.

ROUZINA, I. & BLOOMFIELD, V. A. (1998). DNA bending by small, mobile multivalent cations. *Biophys. J.* **74,** 3152-3164.

SCHULTZ, S. C., SHIELDS, G. C. & STEITZ, T. A. (1991). Crystal Structure of a CAP-DNA Complex: The DNA is Bent by 90°. *Science* **253,** 1001-1007.

SEEMAN, N. C., ROSENBERG, J. M. & RICH, A. (1976). Sequence-Specific Recognition of Double Helical Nucleic Acids by Proteins. *Proc. Natl. Acad. Sci. U.S.A.* **73,** 804-808.

SELSING, E., WELLS, R. D., ALDEN, C. J. & ARNOTT, S. (1979). Bent DNA: visualization of a base-paired and stacked A-B conformational junction. *J. Biol. Chem.* **254,** 5417-5422.

SHUI, X., MCFAIL-ISOM, L., HU, G. G. & WILLIAMS, L. D. (1998a). The B-DNA Dodecamer at High Resolution Reveals a Spine of Water on Sodium. *Biochemistry* **37,** 8341-8355.

SHUI, X., SINES, C., MCFAIL-ISOM, L., VANDERVEER, D. & WILLIAMS, L. D. (1998b). Structure of the potassium form of CGCGAATTCGCG: DNA deformation by electrostatic collapse around inorganic cations. *Biochemistry* **37,** 16877-16887.

SINES, C. C., MCFAIL-ISOM, L., HOWERTON, S. B., VANDERVEER, D. & WILLIAMS, L. D. (2000). Cations mediate B-DNA conformational heterogeneity. *J. Am. Chem. Soc.* **122,** 11048-11056.

SOLER-LOPEZ, M., MALININA, L. & SUBIRANA, J. A. (2000). Solvent organization in an oligonucleotide crystal. The structure of d(GCGAATTCG)$_2$ at atomic resolution. *J. Biol. Chem.* **275,** 23034-44.

STEFL, R., WU, H., RAVINDRANATHAN, S., SKLENAR, V. & FEIGON, J. (2004). DNA A-tract bending in three dimensions: solving the dA4T4 vs. dT4A4 conundrum. *Proc Natl Acad Sci U S A* **101,** 1177-82.

STOFER, E. & LAVERY, R. (1994). Measuring the Geometry of DNA Grooves. *Biopolymers* **34,** 337-346.

STRAUSS, J. K. & MAHER, L. J. (1994). DNA bending by asymmetric phosphate neutralization. *Science* **266,** 1829-1834.

STRAUSS, J. K., PRAKASH, T. P., ROBERTS, C., SWITZER, C. & MAHER, L. J. (1996a). DNA bending by a phantom protein. *Chem. Biol.* **3,** 671-678.

STRAUSS, J. K., ROBERTS, C., NELSON, M. G., SWITZER, C. & MAHER, L. J. (1996b). DNA bending by hexamethylene-tethered ammonium ions. *Proc. Natl. Acad. Sci. U.S.A.* **93,** 9515-9520.

STRAUSS-SOUKUP, J. K. & MAHER, L. J. (1997). DNA bending by GCN4 mutants bearing cationic residues. *Biochemistry* **36,** 10026-10032.

STRAUSS-SOUKUP, J. K. & MAHER, L. J. (1998). Electrostatic effects in DNA bending by GCN4 mutants. *Biochemistry* **37,** 1060-1066.

STRAUSS-SOUKUP, J. K., RODRIGUES, P. D. & MAHER, L. J. (1998). Effect of base composition on DNA bending by phosphate neutralization. *Biophys. Chem.* **72,** 297-306.

STRAUSS-SOUKUP, J. K., VAGHEFI, M. M., HOGREFE, R. I. & MAHER, L. J. (1997). Effects of neutralization pattern and stereochemistry on DNA bending by methylphosphonate substitutions. *Biochemistry* **36,** 8692-8698.

STROUD, R. M., KAY, L. M. & DICKERSON, R. E. (1974). The Structure of Bovine Trypsin: Electron Density Maps of the Inhibited Enzyme at 5 Å and 2.7 Å Resolution. *J. Mol. Biol.* **83,** 185-208.

TERESHKO, V., MINASOV, G. & EGLI, M. (1999). A "hydrat-ion" spine in a B-DNA minor groove. *J. Am. Chem. Soc.* **121,** 3590-3595.

TERESHKO, V., WILDS, C. J., MINASOV, G., PRAKASH, T. P., MAIER, M. A., HOWARD, A., WAWRZAK, Z., MANOHARAN, M. & EGLI, M. (2001). Detection of alkali metal ions in DNA crystals using state-of-the-art X-ray diffraction experiments. *Nucleic Acids Res.* **29,** 1208-1215.

TRIFONOV, E. N. (1991). DNA in profile. *Trends Biochem Sci* **16,** 467-70.

ULANOVSKY, L. E. & TRIFONOV, E. N. (1987). Estimation of wedge components in curved DNA. *Nature* **326,** 720-722.

VAN HOLDE, K. E., SAHASRABUDDHE, C. G. & SHAW, B. R. (1974). A model for particulate structure in chromatin. *Nucleic Acids Res* **1,** 1579-86.

VILLERET, V., HUANG, S., FROMM, H. J. & LIPSCOMB, W. N. (1995). Crystallographic Evidence for the Action of Potassium, Thallium, and Lithium ions on Fructose-1,6-bisphosphatase. *Proc. Natl. Acad. Sci. U.S.A.* **92,** 8916-8920.

WANG, A. H., QUIGLEY, G. J., KOLPAK, F. J., CRAWFORD, J. L., VAN BOOM, J. H., VAN DER MAREL, G. & RICH, A. (1979). Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282,** 680-6.

WANG, A. H., UGHETTO, G., QUIGLEY, G. J. & RICH, A. (1987). Interactions between an Anthracycline Antibiotic and DNA: Molecular Structure of Daunomycin Complexed to d(CpGpTpApCpG) at 1.2 Å resolution. *Biochemistry* **26,** 1152-63.

WATSON, J. D. & CRICK, F. H. (1953). Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171,** 737-8.

WIESBROCK, F. & SCHMIDBAUR, H. (2003a). Complexity of coordinative bonding in thallium(I) anthranilates and salicylates. *Journal of the American Chemical Society* **125,** 3622-3630.

WIESBROCK, F. & SCHMIDBAUR, H. (2003b). Complexity of coordinative bonding in thallium(I) anthranilates and salicylates. *J Am Chem Soc* **125,** 3622-30.

WILLIAMS, L. D., EGLI, M., UGHETTO, G., VAN DER MAREL, G. A., VAN BOOM, J. H., QUIGLEY, G. J., WANG, A. H.-J., RICH, A. & FREDERICK, C. A. (1990). Structure of 11-Deoxydaunomycin Bound to DNA Containing a Phosphorothioate. *J. Mol. Biol.* **215,** 313-320.

WING, R., DREW, H., TAKANO, T., BROKA, C., TAKANA, S., ITAKURA, K. & DICKERSON, R. E. (1980). Crystal structure analysis of a complete turn of B-DNA. *Nature* **287,** 755-758.

WOLK, S., THURMES, W. N., ROSS, W. S., HARDIN, C. C. & TINOCO, I., JR. (1989). Conformational analysis of d(C3G3), a B-family duplex in solution. *Biochemistry* **28,** 2452-2459.

WOODS, K., MCFAIL-ISOM, L., SINES, C. C., HOWERTON, S. B., STEPHENS, R. K. & WILLIAMS, L. D. (2000). Monovalent Cations Sequester within the A-Tract Minor Groove of [d(CGCGAATTCGCG)]2. *J. Am. Chem. Soc.* **122,** 1546-1547.

WULFSBERG, G. (1991). *Principles of Descriptive Inorganic Chemistry*. University Science Books, Sausalito, CA.

XU, Q., DENG, H. & BRAUNLIN, W. H. (1993). Selective localization and rotational immobilization of univalent cations on quadruplex DNA. *Biochemistry* **32,** 13130-7.

XU, Q., JAMPANI, S. R. & BRAUNLIN, W. H. (1993). Rotational dynamics of hexaamminecobalt(III) bound to oligomeric DNA: correlation with cation-induced structural transitions. *Biochemistry* **32,** 11754-60.

XU, Q., SHOEMAKER, R. K. & BRAUNLIN, W. H. (1993). Induction of B-A transitions of deoxyoligonucleotides by multivalent cations in dilute aqueous solution. *Biophys J* **65,** 1039-49.

YOUNG, M. A., JAYARAM, B. & BEVERIDGE, D. L. (1997). Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.* **119,** 59-69.

ZINKEL, S. S. & CROTHERS, D. M. (1987). DNA bend direction by phase sensitive detection. *Nature* **328,** 178-181.